**REPORT DOCUMENTATION PAGE**

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE**
4 Feb 98

**3. REPORT TYPE AND DATES COVERED**

**4. TITLE AND SUBTITLE**
Recursive Fitness

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**
James L. Cook

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Ruprecht-Karls-Universitat

**8. PERFORMING ORGANIZATION REPORT NUMBER**
97-042D

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION AVAILABILITY STATEMENT**
Unlimited Distribution
In Accordance With AFI 35-205/AFIT Sup 1

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

19980311 111

**14. SUBJECT TERMS**

**15. NUMBER OF PAGES**
526

**16. PRICE CODE**

**17. SECURITY CLASSIFICATION OF REPORT**

**18. SECURITY CLASSIFICATION OF THIS PAGE**

**19. SECURITY CLASSIFICATION OF ABSTRACT**

**20. LIMITATION OF ABSTRACT**

## Erklärung

Hiermit erkläre ich, daß ich diese Dissertation selbständig verfaßt, alle wörtlich oder sinngemäß übernommenen Textstellen als solche kenntlich gemacht, andere Quellen und Hilfsmittel als die in der Arbeit genannten nicht benutzt und die Dissertation noch keiner anderen Fakultät vorgelegt habe.

Heidelberg, den 11.7.97 James L Cook

# Recursive Fitness

Dissertation
for the degree of
Doctor of Philosophy
in the
**Faculty of Philosophy**
of the
Ruprecht-Karls-Universität
Heidelberg (Baden-Würrtemberg)

submitted by
© **James L. Cook**
New Mexico, USA

**Heidelberg 1997**

# Acknowledgments

I am grateful to my dissertation advisor, Prof. Dr. Martin Carrier, who responded to an unexpected knock on his door with warmth and enthusiasm. He is literally responsible for every word (actually every *token*, to use philosopher's jargon) on the following pages. Had it not been for his tenacity and ingenuity I might have ended up writing in Latin for a readership of dubious fitness, recursive or otherwise.

I am also indebted to the United States Air Force Institute of Technology and the United States Air Force Academy for their ongoing support, and especially to the Universität-Heidelberg for allowing me to study here.

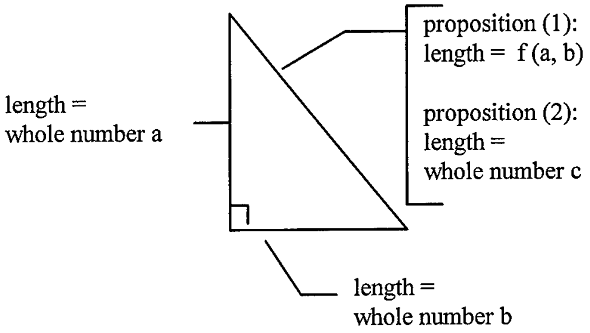
# Table of Contents

# Part One: Foundational Concepts for a Recursive Understanding of Fitness

# Chapter One: Introduction

## 1. Abstract Motivation: Methodically Constructive Observation

The legendary scenario is chilling: A young man sails in the eastern Mediterranean or perhaps on the Aegean along with other members of an esoteric cult. He is the only novice aboard; all the other voyagers are tried and true initiates. His brotherhood lives under arcane rules -- one may not eat beans, for instance -- and dedicates itself to discovering mathematical regularity in nature. The voyage affords time for reflection, and the young monk continues to work on a problem which has occupied him for several days. He scratches right triangles on a board and wonders: How can I show that the length of the hypotenuse can *always* be expressed as some relationship between two whole numbers (proposition 2 below)?



Clearly proposition (1) in the diagram above is true. The length of the hypotenuse must be a function of the lengths of the other two sides, given that the angle between those sides has already been specified (as 90 degrees). But we know now that proposition (2) is seldom true. That is because triangles such as the 3-4-5, in which whole numbers satisfy the equation $a^2 + b^2 = c^2$, are comparatively rare; *most* triangles will not satisfy the equation. But the young monk is not aware that *any* triangles can have an incommensurable hypotenuse. At first he thought that his choice of units must have been at fault whenever a triangle remained intractable, but now he is no longer sure. In fact he is bewildered at the direction in which his investigations seem to lead. For most right triangles he has examined, two "whole" legs can be joined only by a hypotenuse whose length *cannot* be expressed as a relationship between whole numbers. No matter what units of measurement he applies to such intractable triangles, the result is the same.

At last he reports to one of the elder voyagers. Before long all of the senior initiates on board are engaged in animated, even heated, debate. The order's most sacred doctrine has long held that all nature can be revealed as relationships among *whole* numbers (q.v. Coppleston 1953: I: 33 - 36)[1], but the neophyte's discovery seems to contradict this. Long hours pass, and the youth is sent away to the bow of the boat as the elders continue to confer. Their voices die from derisive shouts of disbelief and boasts that a solution has been found to conspiratorial whispers, and not infrequently the young monk notices a malignant glance cast in his direction. As darkness falls, the elders come forward. The senior member issues a curt command. Used to unquestioning obedience, the neophyte submits and is soon bound hand and foot. Another order is given and the youth is thrown overboard into deep water, many miles from the nearest coast. All on board are sworn to secrecy -- not so much to conceal the murder (or *execution* for the good of the order and of mankind, as the elder monks would have seen it). Rather, the vow of silence was intended to hide the apparent existence of incommensurability.

True or not, this scenario is consistent with a legend about the Pythagorean brotherhood. The philosopher and mathematician Pythagoras is said to have founded the order sometime during the mid- to late-6th century B.C. There are different versions of how incommensurability was discovered and of what transpired in the find's aftermath. Some stories are mild, claiming only that the cult attempted to "suppress" the existence of irrationals (Bergamini 1980: 43 - 44), but the tale of the discovery at sea and the subsequent drowning of the discoverer lives on (Alioto 1987: 40; Russell 1959: 22). For anyone interested in the relationship between method and quiddity, it is worthwhile to consider the storied incident and the doctrine which served as its background as indicators of the way in which previously unseen quiddity can arise from method.

Sometimes the discovery of new methods of analysis can have dramatic and long-lasting impacts on the ontology of a scientific discipline and of observation itself. The Pythagoreans are believed to have equated being itself with number, but they must have realized that when we listen to a musical scale or watch the sun set, we do *not* perceive *in terms of number*. We do not say to ourselves, for instance, "How clever of that composer, to have used 15/16 of the previous tone as the lowest note in the next passage!" Nor could a Pythagorean brother have glanced at something like a $1\text{-}1\text{-}\sqrt{2}$ triangle and thought "Oh, oh...this looks like trouble!" Rather, the brotherhood at Cortona observed number in nature through the aid of numerical *methods*. The drowning legend underscores this fact. Things were going along swimmingly for the Pythagoreans -- regularities involving *whole* numbers were being found in music and throughout nature -- until the problem of incommensurability was encountered and the doctrine that "all is (whole) number" began to founder. It was not as though members of the cult had never seen a triangle comprising incommensurable legs before. Indeed they must have, but neither they nor anyone else could have recognized the fact without first wielding certain numerical methods.

Before going further here, it would be well to clear up a point of ontology. We can read the Pythagorean doctrine as claiming that number is a kind of conceptual overlay which the observer places upon the "nature" available through our senses, or we can say that a realist's independent nature actually is essentially numerical and would be so whether observers schematized it with the aid of numbers and numerical methods or not. By this second account, numerical methods are tools of what we might call *discovery* rather than *construction*. For our purposes there is no need to determine which is the proper interpretation of Pythagorean doctrine, nor must we decide the more general question -- whether some version of realism is correct or not. Here it is enough to find a means of distinguishing between perception and that which is perceived, and such means -- for instance Whitehead's 1919 conception of "thinking 'homogeneously' about nature," von Ditfurth's 1981 "hypothetical realism," or Putnam's 1990 "internal realism" -- need not be vehemently partisan in the ontological arena.[2] Having found such a method, we will then have at least some warrant for speaking of quantifying nature rather than merely our own thoughts about nature. For our purposes it will not matter whether the distinction is merely convenient, that is, whether it is the case that when we think about an "external" nature we are in fact still "thinking about thought" (in Whitehead's phrase).[3]

Returning to the Pythagoreans, then, we can appraise them as observers either *discovering* or *constructing* nature. Which identity we choose for them is not so important; what matters to us is that they did whatever it is they did with the aid of numerical methods. Under either a realist or anti-realist account of the Pythagorean doctrine, the deepest secrets of reality as number are not revealed to the casual observer, even given the assumption that quantity, proportion and geometric form constitute the essence of all things. All may be number in some sense, but the interpretation of that generality depends upon very specific, often unobvious algorithms. Whatever is *out there* -- in a realist's nature or in the mental realm which is being thought about -- reveals itself only through methods of analysis which blur the boundary between observation of the existent and manipulation of the observed. This generality is applicable much beyond Pythagoreanism, of course. Almost 2500 years after the brotherhood flourished, another monk spent much of his time carefully quantifying natural phenomena. Fortunately Mendel's cult did not proscribe legumes of any kind, and his work with peas demonstrated an unobvious numerical regularity in the natural world. As was true of any number of earlier discoveries (such as the Pythagorean "quantification" of musical scales), Mendel's findings cannot be *naively* observed; instead they must be *calculated.*

Consistent with but not explicitly contained in the Pythagorean outlook is the belief that possession of certain algorithms makes us *willing* as well as able to perceive things which we might otherwise treat as merely a part of the background noise -- of the "blooming, buzzing confusion," to borrow William James's phrase -- in the world around us. An engineer in antiquity might never have taken the area under a given curve into account in his calculations, no matter how relevant that "object" would seem to his modern colleagues. Problems the engineer would have ignored as insoluble are now amenable to analysis using various methods discovered and described by Newton between about 1665 and 1704[4]: integration through infinite series, fluxions, as well as prime and ultimate ratios (Smith 1959: 613 - 618). The very meaning of a "fluxion" as a "velocity of motion, or increment" (ibid. 614) could be said to emerge from the *method* used to extract such an entity from its quantitative environment. In the absence of a general method, one which works somewhat independently of the context and purpose of observation, the observer might well be restricted to conceiving of speed and acceleration only (first and second derivatives, we might say), while the abstract notion of a fluxion (something which would accommodate third and higher derivatives) would not exist. One caution is in order at this point. Numerical analysis may create certain entities, but that does not entail that the process of instantiation-by-method extends further. If numerical methods are used to observe and characterize the force of gravity, for instance, they do not thereby imply the existence of qualia which cause gravity to exist in the way described (q.v. Koyré's analysis of Newton's famous "*Hypotheses non fingo*" -- 1957: 228 - 233).

Other examples of how method "creates" the objects it manipulates are more obvious. Concluding his explanation of decimal fractions -- the first ever published -- Simon Stevin of Bruges (c. 1548 - c. 1620) urged their use for all weights and measures. By that he clearly meant that new standards should be *created*, for he speaks of a decimal future: "If this [system] is not put into operation as soon as we might wish, we have the consolation that it will be of use to posterity, for it is certain that if men of the future are like men of the past, they will not always be neglectful of a thing of such great value" (1634: 34 in Smith 1959).

It could almost be said that "objects" such as incommensurable triangles, derivatives and decimal fractions do not exist -- or perhaps more precisely and realistically, that they have not been brought into perceived existence -- until a method for their analysis and use is known. But of course the process can work in reverse, too: ontologies can crumble and blow away as new methods and new discoveries supplant older models and their referents. With the advent of the calculus, a number of older methods (and the world-views associated with them) ceased to be of interest, no matter how earnestly and ingeniously they had been investigated in the past. The same is true in other domains. We know that Newton was an ardent alchemist, performing "untold scores of alchemical experiments" and producing "alchemical notes and manuscripts [which] have been conservatively estimated at 1,000,000 [words]" (Christianson 1984: 110, 203; cf. Weeks and James 1995: 92, 113 - 114), but his insights in that field do not contribute to contemporary science in the way that others of his discoveries do. Once a certain chemistry came to prevail, it no longer much mattered what was said about, say, transforming other metals into gold, nor did it matter who said it. Even theoretical constructs which "work" given an extant phenomenology can be replaced by other models, and when that happens, the ontologies associated with earlier models lose meaning and interest. Phlogiston had already been used as a "combustibility principle" by the time Cavendish equated it with hydrogen and used it with apparent success in his own chemical architectonics (Carrier 1992), but phlogiston *per se* is not of burning interest in modern chemistry textbooks.

One has only to consider these few examples (and the list is easily expandable) to see a pattern: the notion that number and geometric form as the essence of nature (whether they are observer-independent aspects of an external nature, conceptual overlays placed upon an external world, or generalities about thought thinking of itself) are not naively perceivable but instead emerge through unobvious manipulations; a visceral aversion to something as emotionally neutral (from a late 20th-century scientific standpoint) as irrational numbers; the inference of an inheritance principle based on numerical regularity; the ability to *perceive* an entity such as a third derivative in nature because we have a *method* of distilling it from what we see; the observation of natural phenomena in terms of one set of units rather than another for the sake of effortless manipulation of nature's "measure"; the supplanting of one chemical ontology with another for the sake of the internal coherence of a discipline's observational and theoretical results. All these instances argue not just for the theory-ladenness of observation but also for some close relation (which we have not specified) between observation and manipulation. We might further speculate that the questions we bring to the workplace of observation are suggested at least in part by our available methods of manipulating things like numbers and geometric forms. (One of the most radical versions of this thesis was framed in the context of linguistics. A strong version of what is called the Whorfian hypothesis says that we cannot even think about what we cannot speak about, or expressed generally, that language determines thought; q.v. Lyons 1981: 261 - 2,.) Looking at the questioning, observational stance as a function of our extant techniques of schematizing and manipulating our sense data, we may see Dewey's generalization manifested in the examples already offered:

> Old ideas give way slowly; for they are more than abstract logical forms and categories. They are habits, predispositions, deeply engrained attitudes of aversion and preference. Moreover, the conviction persists -- though history shows it to be a

> hallucination -- that all the questions that the human mind has asked are questions that can be answered in terms of the alternatives that the questions themselves present. But in fact intellectual progress usually occurs through sheer abandonment of questions together with both of the alternatives they assume -- an abandonment that results from their decreasing vitality and a change of urgent interest. We do not solve them: we get over them. Old questions are solved by disappearing, evaporating, while new questions corresponding to the changed attitude of endeavor and preference take their place. (Dewey 1910: 19)

For Dewey, "the scientific revolution that found its climax in the 'Origin of Species'" is "[d]oubtless the greatest dissolvent in contemporary thought of old questions" (ibid.). And that sentiment brings us close to our present focus.

Abstractly, this dissertation is motivated by something like Dewey's abstract insight, only here the new questions have a definite wellspring -- what we can call *method* or *technique* or *algorithm*, terms which amount to a means of what we can call *constructive observation.* Means of manipulating the "raw data" of perception may reveal novel ways of posing questions about an area of scientific interest. In turn, questions asked from a new perspective may be more fruitful than queries which were meaningful in older contexts. We can hope that the answers will allow us to see or hear the bloom or the buzz or whatever else is under investigation more clearly and in a way that jibes more seamlessly with the rest of our conceptual commitments.

## 2. Concrete Focal Point: Fitness evaluated as recursive

Our specific focus in this dissertation will be fitness as the concept is used in evolutionary biology (and also, it will be argued, in "hindsight-constructive" disciplines such as historical linguistics in general). The "new" method of analysis -- specifically the new means of reevaluating fitness and its use in evolutionary biology -- is recursion. More precisely, it is the description of fitness as a recursive function or object which is new here, for certainly recursion itself cannot be called a novel method in domains such as computer science. In fact many modern presentations find rather old inventions (e.g., Fibonacci numbers, dating from about 1202; q.v. Hofstadter 1979: 135) to be productive means of explaining what recursion is and simultaneously demonstrating how recursive algorithms might be "coded" in an explicitly self-referential way (Wirth 1986: 119). If it would be incorrect to say that recursion is a new discovery, however, it is nonetheless true that the sudden ubiquity of computers has made recursion a practical method for solving certain problems and defining particular situations which would not have been amenable to that approach a fairly short time ago.

The salient term here is *practical*. Recursion as roughly a genus of self-referential algorithms and data structures may have been conceived even before Fibonacci, but sometimes discoveries must be put on the shelf for a time before they pay-off in the practical sphere. In 1991 a congressman asked a subcommittee in the U.S. House of Representatives to authorize a coin commemorating the 500th anniversary of Columbus's "discovery" of the New World. Another congressman, who represented a large constituency of Scandinavian ancestry, suggested that Leif Ericson rather than Columbus may have been the first to make landfall in the Americas. "'Well,' [Representative] Annunzio told [Representative] McCandless and the audience, 'when Columbus discovered America, it *stayed* discovered'" (Seppy 1994: 88). So it is with recursion (and perhaps other methods depending on efficient automation): recursion is a "New World" which one could have visited only with great difficulty until a means existed of rapidly repeating many iterations of what is essentially the same, self-referential algorithm. Given that new-found ability, recursion can now "stay discovered," too, meaning we can use it frequently and easily in areas where once it would have been impracticable.

Why we would want to use such a method is probably clear based on two observations. First, scientific models strike compromises between static and dynamic ontologies. What appears as a single property of an element in a model system may be in constant flux in the *real* system, that is, in the one modeled. A certain model may identify fitness, for instance, as a relatively constant property of a certain organism in a given selective environment, but even proponents of that model may recognize that fitness, understood as the organism's propensity to possess a certain longevity and degree of reproductive success, is in constant flux. Secondly, historical inquiries occupy themselves in large measure with "filling in the gaps" between data points. Methods (such as recursion) which are essentially cumulative lend themselves to such an endeavor: we take a known state of a particular organism, perhaps represented by a fossil, and then we reason through millennia which offer no empirical data points by speculating how the organism would have been modified at each step *based upon its status in the previous iteration of analysis*. We judge our theorizing to have been successful if we can build conceivable bridges between data points -- if we can explain the origin of extant species as the descendants of organisms represented by extant paleontological evidence, for example. An analogous case might be the attempt to explain how a sum of money in a bank account in 1950 "became" the present sum in that account. Assuming we had one figure for each of the intervening decades as well -- the 60s, 70s, and 80s -- the challenge would be to build a model of accumulation (or decrease) to explain the progress of sums. The easiest way would be to introduce an external agent who makes the appropriate deposit or withdrawal before every extant data point. Instead of *deus ex machina* (the god, raised by a crane above the stage in some Greek and Roman dramas, who decides the play's outcome) the decisive agent at the bank is an *emptor ex machina* (only the machine in question is not a crane but rather an automatic teller). More challenging is to find an algorithm which manipulates the initial amount of money in such a way that each of the data points is "predicted" (albeit in hindsight) and without an external agent such as the *deus* or the *emptor*. An algorithm calculating compound interest might do the trick, for example, by taking the initial value, manipulating it, then taking the result of that manipulation, manipulating it in the same fashion as before, etc. Such an algorithm *need* not be defined recursively, but as we will see below, it sometimes can be. Moreover, some algorithms (such as the one calculating compound interest as just described) seem to lend themselves best to recursive expression. I take that as a fact, but I admit that I cannot prove it in any rigorous way. Perhaps it must be accepted as an aesthetic matter. (Consider how an algorithm for calulating compound interest *non*-recursively would compare to the essentially recursive one described above.)

More generally, imagine that we are trying to build a model of a dynamic system S. The system consists of elements $\{E_1, E_2, ...\}$, each of which possesses a certain value for a general quality Q. A given element $E_i$ is then "more Q," "less Q," or "just as Q" as another element $E_n$. Forces $\{F_1, F_2, ...\}$ are responsible for the interactions of the elements. Each $Q_i$ corresponding to a particular element $E_i$ of the system varies depending on the makeup of the system as a whole, including the Q-values of the other elements. Moreover, suppose we can adequately evaluate the $Q_i$ of a given element only across time and not as a mere "snapshot," either because we cannot accommodate the agency of the operative forces unless we know initial conditions or else because we do not know enough about the dynamics of the system in general. If the elements in question were, say, bodies in space whose masses we knew but whose velocities were unknown, if only one force (e.g., gravity) were operative, and if the quality in question were velocity, then we could not say what the velocities of the objects were at a given moment. The reason for our inability is that gravitational forces affect *already existing* motions; we can express velocity formally as a function of gravitational force, but we cannot say what particular velocity a moving object has unless we know its initial or previous velocity. Moment by moment each Q-value changes as other aspects of the system change, but each $Q_i$ is also dependent on the earlier Q-states of the same element. Thus each $Q_i$ at a time $t_m$ is a substrate for later Q-values at $t_n$ ($n > m$). If we wanted to evaluate $Q_i$ numerically, we could then use a self-referential expression -- some variation of $Q(t_n) = f(Q(t_m))$.

One of the primary contentions of this dissertation is that fitness lends itself to such a recursive definition. Moreover, it will be argued that "reductions" (micro- or otherwise) of fitness values to concrete circumstances do not exist except perhaps as a recursive self-reference (if that is not too great a strectch of nomenclature). To return to the analogy of moving bodies, we have never known an initial number corresponding to a fitness value in the way that we can know the initial velocity of a body in motion. To complicate the matter further, in the latter case we can directly observe and measure the current state of the quantifiable characteristic in question if we happen not to know its initial value; but the same is not true in the case of fitness. When it comes to reckoning fitness, we build a ship which is already under weigh (to use Neurath's metaphor, quoted in Quine 1969: xii). We can assemble data, true enough, but the data are relevant only when we interpret them with reference to past states of the same quality with respect to which the data are gathered.

The primary question to be asked about the notion of fitness in evolutionary biology has traditionally taken what might be considered a natural form: What is it? But the focal point of that question, the "it," is in general not approached directly, not in the way we would approach the question, say, What is a monkey? Rather, the question "What is it?" is translated into "What does it *do*?" in such a way that the "it" is subordinated to its effects. In itself the translation is not regrettable, and one would be hard pressed to find a way of avoiding it. But the translation becomes dangerous when we forget that it has occurred, when we think that answers about fitness's effects are answers about fitness itself.

There is another danger when it comes to a traditional understanding of fitness. Very specific fitnesses are routinely *calculated* in various ways by practicing biologists, while perhaps the best-known abstract analysis of the nature of fitness calls it an "intrinsic property" which is essentially a "propensity" (Mills and Beatty 1979). On the other hand, biologists have very general means of describing the method of calculating fitness, and the vogue among philosophers of science is to treat fitness as multiply-realizable (q.v., e.g., Sober 1984a, 1993; Rosenberg 1985). Two not wholly compatible natures are thus assigned to fitness: one moment and for one purpose it is said to be a specific value, but another analytic goal drives fitness into a wholly different category. This second identity seems very like a function -- some general pattern or method for which an argument or arguments must be plugged in before a specific value is yielded. The two viewpoints might be characterized as opposite ends of the spectrum of causality. One sees fitness as a cause of phenomena such as differential longevities and reproductive successes, while the other perspective takes fitness to be the effect of concrete circumstances. But the knowledge base necessary to evaluate fitness from one perspective derives from the opposite perspective.

At this point the reader may object that all of this amounts to hair-splitting (and fear that 500-some pages of similar bisections await). Granted, the term "fitness" is sometimes applied to specific numbers and other times to more general expressions (the objection will go), but that is a matter of convenience and certainly causes no confusion. However, let us follow Frege (1891) for a moment and reconsider. If we admit a fundamental difference between *Funktion* and *Gegenstand* (or more generally between *Sinn* and *Bedeutung*) in the abstract, then surely we confuse the two at our metaphysical peril when we employ specific propositions. In fact, as we will see, a confusion traceable to the flouting of the distinction between fitness as algorithm and fitness as argument has led to misunderstandings about what it means to "reduce" propositions involving fitness to claims about specific circumstances (e.g., in Weber 1996).

This dissertation will try to separate the respective senses of fitness as function and as value, while offering what I take to be a new and fruitful understanding of what *kind* of function fitness is. I argue that a useful (but not necessary) paradigm for fitness as function and argument should be *recursion*. Previous perspectives have treated fitness as a more or less static entity, and this is true of fitness as algorithm (as the formulae in textbooks on evolutionary biology will attest; cf. Futuyma 1986: 151, 251) and as argument (as phrases such as "intrinsic property" imply; q.v. Mills and Beatty 1979). But a look at the work of practicing evolutionary biologists makes clear what seems to be almost an analytic proposition -- that the interaction between changeable organism and variable environment is bound to be variable itself. Doubtless such an observation will again evoke a "common-sense" objection: Of course there is a dynamic involved, but we have to find ways of "freezing the action" for practical purposes, as when judges at a foot race sometimes revert to a freeze-frame of the finish line in a particularly close race. If we did not do so, the skeptic will argue, much of the explanatory and predictive power of evolutionary biology would be lost.

However, that objection should not be taken as a justification of the status quo, but rather as a statement of the reason why we should find the *best* possible formal paradignm for our ways of circumscribing and summarizing observations of evolutionary phenomena. When we speak loosely of specific numerical values (*Gegenstände* rather than *Funktionen* in Frege's terms), I believe we do so with the understanding that these quantities play a role in an *ongoing* recursive function. That is far different from saying that the values *are* the function (essentially Mills and Beatty's gripe with non-propensity interpretations). But even propensities have a certain static character when we treat them as temporally enduring, intrinsic properties of organisms and taxa. A more detailed examination of received interpretations of fitness will make it clear that such static description of what is an essentially dynamic, self-referential concept tends to confuse the interlocutors in two key debates -- the debate over the reducibility of fitness and the question of what is really the unit of selection. By treating fitness as a recursive function yielding ephemeral "objects" which are also to be understood recursively, significant progress can be made toward resolving these issues.

It is important to stress, however, that this dissertation is not another hairpin in a series of linguistic turns. The intention here is not to argue that by carefully and correctly defining fitness (the "correct" answer in this case meaning a *recursive* definition), all problems magically evaporate. On the contrary, it should be emphasized at the outset that I take recursive fitness to be a wholly corrigible concept. Some corner of some discipline -- of ecology, say -- may be bursting with examples of which I am unaware and which are totally intractable when approached recursively. Further, a better model of analysis may be right around the corner or even under our noses. The claim here is simply that viewing fitness as a recursive function jibes well with what biologists do while helping to avoid some of the logical pitfalls plaguing philosophers of science. This should be no surprise, given the examples which began this introduction and Dewey's sentiment about the way in which questions in scientific disciplines arise and fall. It may be that in ages and environments where computers (conceived as machines which handle stacks efficiently) are absent or sparse, recursion as a method simply could not have become sufficiently widespread to serve as a model of analysis. That is not to say that recursion is dependent upon computers. As we have seen, introductions to recursive method frequently use pre-Digital Computer Age "objects" such as factorials and Fibonacci sequences as examples, not to mention phenomena known through everyday experience (e.g., Russian dolls) as metaphorical representations (Hofstadter 1979: 127; Wirth 1986: 135 - 6). But recursion became a methodological "headline" only with the dramatic spread of digital computer technology. No wonder, then, that fitness has not been analyzed as a recursive phenomenon before now (assuming that the approach is at all fruitful).

Recursion can be viewed abstractly as the calculus of our computer-laden age, as a continuation of the methodology pioneered by Newton and Leibniz for analyzing complex scenarios. One way of expressing the *modus operandi* of the calculus is "shrinking and multiplying": by *shrinking* the context of analysis, large-scale changes can be atomized and described in detail; by *multiplying* the number of times we observe such carefully circumscribed fields of examination, complicated things which change continuously (such as areas under some curves) can be precisely reckoned. Recursion is an especially apt means of shrinking and multiplying, too: any size of incrementation and any number of iterations are theoretically possible.

Narrowing the scope of individual observations presupposes conceiving and identifying constituent parts of a macro-phenomenon, whether in an empirical or a formal context. There was no warrant to speak of the components of white light until Newton demonstrated its composite character through a series of experiments carried out between 1664 and 1666. Similarly, the notion of a first derivative, dy/dx, makes sense only if we know a method to distill the smallest unit of observation from a curve. Here, the stress on method is not meant to carry Percy Bridgman's or anyone else's banner in the debate over how meaning is constituted. Certainly we might conceive of a vanishingly small portion of a curve without knowing a method for quantifying such an entity. What is important for our present purposes is that *some* concepts are operationally defined: *sometimes* manipulation yields insight into quiddity. Of course it can always be argued that the idea of a more fundamental unit precedes the inclination to tinker in many if not most cases of scientific experiment. Would Newton have bothered with prisms if the truth about light had not already dawned on him (if only in the form of a suspicion -- the kind which scientists are wont to call "hypotheses" and cartoonists represent with glowing light bulbs above a character's head)? Perhaps not, but presumably the experiments yielded something more definite than whatever he might have envisioned beforehand.

What more can be said about the entities uncovered through manipulation? First, algorithms sometimes yield things which are at once emergent and primitive. If that sounds paradoxical, perhaps we should turn back to our first derivative. Units such as dy/dx are monad-like; they are atomic, or indivisible, in the formal sense that a given unit has counterparts which differ from it in no way except their outer context. A first derivative of 3.78 on one curve is the same as a first derivative of 3.78 on a wholly different curve if the criterion of comparison is restricted to quantity alone. In fact, there is not even a *qualitative* difference between the two derivatives if they are viewed apart from their respective "environments." Yet in so far as a first derivative *qua* ratio of limits is reckoned from other quantities -- the changes in ordinate and abscissa (if we're talking about two-space) which form the respective parts of the ratio -- the derivative must be *emergent* from its parts. This brings us to a second reflection about the distillate of algorithmic manipulation. Although something like a first derivative can be seen as emergent from other entities, it is not necessarily reducible to them *operationally*. That is because some manipulations are what we might call "one-way." We can say that a first derivative in two-space is a ratio of two vanishingly small changes, expressed as quantities, but the limits involved are mute on the question of what the components of the changes are. To repeat, a specific derivative as quantity might accurately describe a point on any number of curves.

Again speaking abstractly (even metaphorically), the debate over the units of selection in evolutionary biology can be viewed as an attempt to restrict the context of observation -- to look at $dy_i/dx_i$, so to speak, rather than at the whole curve. This is not to deny the obvious: that the units of selection problem was and continues to be motivated by other goals and issues. One such goal is lessening the tension between apparently altruistic behavior on the one hand and the "Every organism for itself!" mentality which underlies Darwinism on the other hand. (Under the category of altruistic behavior we can include the role of sterile servants which, among organisms such as termites, ants, bees, and wasps, help their fertile siblings pass on genes that both categories, fertile and sterile, share.) Another issue motivating the units of selection debate is the phenomenon of "abnormal" sex ratios which seems to violate Fisher's sense of why a one-to-one ratio is generally desirable. But the units debate can also be seen as a quest to find a *manageable* context of analysis which serves to validate the principle of evolution by natural selection, that is, by the weeding-out of traits relatively ill-suited to an environment in favor of other traits comparatively better-suited. It will be argued that that "shrinking" of the context of analysis functions primarily as a *recursive* algorithm whenever we talk about fitness.

The method of analysis which makes the shrinking or atomizing function of the calculus possible depends upon the concepts of limit and sequentiality. The same can be said of the self-referential aspect of recursion. Describing the essence of a function which is defined in terms of itself presupposes the ability to understand the way a progression of "self-like" units add up to the self, itself. The awkwardness of such a sentence reflects the difficulty of conceptually separating function and argument (returning to Frege's terminology) when both share the same designator, as in a procedure (function) which produces a Fibonacci sequence by beginning with one argument and then calling itself (qua function) with itself as argument. What is the thing called -- function or argument? Something similar seems to happen in the units of selection debate, when we recognize that a given "level" (say, the individual as a whole) is a function of another level (the genotype or a combination of genotypes).

If the discussion over the units of selection corresponds to the "context-shrinking" function of the calculus in our analogy, what corresponds to the "observation-multiplying" function? Practical users of recursion -- computer scientists, for instance -- can answer at once: A recursive procedure can call itself an unlimited number of times (although that is not desirable from a practical point of view). In the case of recursive functions, quiddity in the sense of meaning or specific value (*Bedeutung* or *Gegenstand* in Frege's terms) emerges from repetition. That is perhaps not so earth-shaking; after all, the same could be said of an iterative function (e.g., a DO WHILE loop in a computer program). But the recursive function as triad (function, argument, *and* value) can be said to emerge only in its own execution. The flow of control in an iterative procedure is subject to various conditions, but all of the constituent parts can be recognized and described in isolation from the others.

There are three criteria by which recursive fitness (or any other interpretation of the concept) must succeed or fail, criteria we can state in the form of questions: Is recursive fitness logically consistent? Does it account for the way in which practicing evolutionary biologists use the term? Is it "productive" (that is, does it contribute to our understanding of nature, vague though that may sound)? Below we will see justifications for answering all three questions in the affirmative. I argue (or at least indicate) (1) that fitness can be understood as a recursive function without logical contradiction, (2) that biologists use fitness (and perhaps many concepts whose meaning is "filled-in" through a hypothetico-deductive method) as a recursive function, and finally (3) that a recursive interpretation of fitness contributes to the solution of several long-standing issues.

The criterion of bare logical consistency is difficult to separate from the productivity yardstick, because the quest for consistency gains urgency when philosophically problematic aspects of evolutionary biology are confronted. From that collision of concepts we learn, for example, whether recursive fitness can accommodate the agency of chance -- one of the primary factors which drove Mills and Beatty (1979) to develop their propensity interpretation of fitness. We will want to know, as well, whether recursive fitness belongs to metaphorically selfish genes, to individuals, or to some other "level" of life (a question which exercised Dawkins, among others), and whether recursive fitness is amenable to reduction of any kind, e.g., the microreduction which Weber (1996) thinks can sometimes be accomplished under a more traditional interpretation of fitness. We may further ask whether a recursive reading of fitness will help resolve the debate over punctuated equilibrium as an alternative to a gradualist account of evolution's course -- an issue which still stirs fervent commentary (e.g., in Dennett 1995). As for the aptness of recursive fitness in the endeavors of working biologists, examples throughout the presentation will be used to imply (though perhaps not to rigorously argue) that biologists employ an essentially recursive means of evaluating the extent to which organisms have adapted to given environments. This should not be unexpected given the hypothetico-deductive template governing experimental and historical biology, but paradoxically, Mills and Beatty's 1979 propensity interpretation seems to be about as close as philosophers of science have come to seeing fitness as what we might call a progressive *function*. Because the notion of function may offend some readers' intuitive sense that fitness is a *property* of the organism and so should have an other than algorithmic kind of quiddity, it will be shown that fitness may also be conceived as what a computer scientist would call a *recursive data structure* -- a "thing," in other words, but one which is recursively defined. Here again the productivity criterion enters the discussion naturally. After all, even if recursive fitness passes the tests of logical coherence and consistence with practical usage, it would seem pointless to bother expounding the concept unless earlier interpretations are in some ways less successful within the context of evolutionary biology as a whole. The fecundity of the concept can only be revealed in the struggle to solve major problems.

To summarize, the results to be had by confronting problems of evolutionary biology with a recursive model of fitness are relevant to the following issues:

### (1) The "circularity problem"

In broad terms, this is the allegation that propositions juxtaposing "survival" (and like concepts such as "longevity" and "reproductive success") with "fitness" in discussions of evolution are inevitably tautologous or analytic. Under a recursive concept of fitness, this problem largely evaporates. Although this may sound like verbal sleight of hand reminiscent of Logical Positivism ("just get your definitions straight and logical problems disappear"), a moment's reflection may suffice to see why calling a "fitness statement" circular *in a recursive sense* is actually a kind of affirmation that a first logical wicket has been successfully passed. If we treat fitness as a recursive algorithm yielding a Boolean output (as fit as or not as fit as the result of the last iteration), it should not be surprising to find a marked similarity in the expression of subject and predicate. That kind of "circularity" might seem vacuous if the self-referential expression were static; but recursive algorithms are best conceived as dynamic, as generating interesting output despite their formal circularity.

Perhaps no verbal weapon wielded by philosophers is deadlier than the word "circular." That is because this description labels a bad argument the result of sloppy thinking rather than something more easily forgiven, such as misperception. A critic may deem a philosopher's conclusions to be false, and certainly that is no compliment. But at least an allegation of plain falsity allows for the possibility (if the critic wishes to be charitable) that the miscreant reasoned well yet went astray because of mistaken assumptions. Philosophers are by and large children of their Greek forebears in believing that every undertaking has its supreme virtue; in philosophy, reasoning well within the scope of one's assumptions is at least as important as assuming what should be assumed. It is hard to know whether one of these two aspects of argumentation -- getting one's premises right or constructing good arguments on the basis of those premises -- is more important than the other to practitioners in disciplines such as biology. Richard Dawkins *qua* ethologist castigates certain of his peers in biology rather severely when he charges that "[i]t wasn't their probability theory they got wrong [in making an allegedly fallacious argument], but their biological assumptions" (1982: 191). Among philosophers it is equally harsh to accuse an argument of being circular -- although the meaning of that charge is often unclear. Or perhaps it is even worse to accuse a philosopher of bad reasoning given assumptions than bad judgment in selecting premises. That is because the allegation of bad reasoning gets at the basic skill in the discipline. One can always change one's assumptions as quickly as socks or underwear, but reasoning ability is the product of years of training. Call a philosopher's perception of the world myopic and she may in time pardon you or even agree with you, but charge that her pet argument is circular and you've made an enemy for life.

No wonder, then, that many philosophers bristle or run for their lives and reputations when the charge of circularity is leveled against some favorite argument or theory. This is as true of philosophers whose focus is biology as of practitioners in other aspects of the discipline. In their bailiwick, the allegation of circularity has focused on the concept of fitness (sometimes called adaptedness, as in Brandon 1987), though one senses that some critics of evolution stumbled across the concept of fitness and its alleged circularity during a more of less blind search for *any* point of attack. A practicing biologist may yawn at the allegation that the concept of fitness is frequently bandied about in circular statements. For the philosopher of biology, on the other hand, such a charge evokes a fight or flight reaction. Those philosophers who have taken up the defense have done their share of both fleeing and fighting. The *flight* has generally taken the form of acknowledgment that some propositions of evolutionary biology are circular in the sense of being somehow empty of empirical content. (The precise meaning of circularity and "emptiness" will be investigated in some detail below.) Such retreats are often followed by a defense of evolutionary biology at large. The tactic is to desert abstract statements about fitness while defending the scientific integrity of evolutionary biology's general program (which amounts to claiming that many of the propositions which interest evolutionists are also testable, i.e., falsifiable à la Popper). The *fight* on behalf of fitness has been waged by translating statements involving the concept into *probabilistic* relationships between specific properties of some biological entities (e.g., genes, individuals, groups of individuals) and the survival of the same or different entities.

My primary purpose in appealing to recursion in this context is to defend circularity in the theory of evolution as a whole by defending the use of the concept of fitness in specific propositions. To make my case it will be necessary to point out the weakness in common responses to the charge that statements about fitness are often circular while arguing for the explanatory power of the concept in many propositions. In other words, the argument is that not all genera of circularity entail vacuity. Thus much will be said about past views of fitness, but not for the purpose of sketching a history of the concept. Criticizing past views is simply a means to a present end: defending propositions involving fitness even though they may legitimately be called circular. If there is a single key to this notion of an innocuous, even enlightening circularity in statements involving fitness or adaptedness, it is the distinction between a finished argument and an ongoing investigation. It will be suggested that propositions of the general form "The fittest survive" can indeed be viewed as interesting claims about the reality framed within a theory. That should be no surprise. But it will be argued that such propositions -- we will call them "ratchet" claims or functions in chapter 9 -- are not themselves circular in a vacuous sense. What is circular is a *process*[5] of argumentation within which a given ratchet claim serves first as a tentative, specific proposition, then as a confirmed bit of evidence for the same claim viewed abstractly. In other words, a given proposition will be seen to have a dual nature, specific and abstract, so that if one pays attention only to words (or perhaps a combination of words and other symbols, such as mathematical ones) it will appear that a proposition refers to and tests itself. Ratchet claims figure in theories which are circular in the sense that they are *recursive*.

To help make this concept of non-vacuous circularity easier to swallow, we might use the modifier "circling" rather than "circular" to emphasize that fitness plays its conceptual role not in something finished, which a substantive such as "argument" already tends to connote, but rather in the course of something which remains in motion. We could call this moving thing a *process*, and although the particular process called recursive fitness is rhetorical (like an argument), it is not finished and will never be, though its course may be modified with decreasing frequency. The importance of distinguishing between circular and circling, between vacuous and non-vacuous sorts of self-referential propositions, emerges in the claim that defenders of evolutionary biology have been running away from the charge of circularity when in fact they should have admitted and even touted the circular character of their discipline's arguments involving fitness. It should be noted that while my view bears generic similarity to theories which defend the concept of fitness (Mills and Beatty 1979) as opposed to those which surrender many propositions involving the concept (Sober 1984b, 1993), my strategy has less to do with reinterpreting fitness itself and more with explaining how any number of concepts can be employed recursively. In short, my method is to demonstrate how a specific kind of circularity is productive and even necessary, rather than to reinterpret fitness in particular.

At this point I would expect the reader's reaction to be skeptical -- suspicious that I now intend to beat a dead horse for five hundred and some odd pages or else that I'm going to saddle the corpse and try riding it for the same distance. The first group of skeptics will yawn in anticipation of a long and pointless argument that science's hypothetico-deductive method (or some variant thereof) progresses through multiple tests of a given proposition. Something much more subtle and (I hope) powerful awaits these doubters; for now, suffice it to say that if my arguments were so time-worn, then one would expect some at least cursory treatment of recursive circularity in the best-known discussions of the circular fitness problem. In fact, those who seek to rescue fitness have done their best to distance themselves from the notion of circularity rather than to ask whether there may be different sorts of circularity and to embrace some subset of the variety as positively useful and common in science. The second group of skeptics may have signed up to the position, best developed by Sober in 1984 and 1993, that certain propositions involving fitness are in fact circular and are, more importantly, essentially empty. (By this account, the salvation of evolutionary biology is that there's more to it than such statements.) My intention is to accept the allegation of circularity but reject that of emptiness. There is no point in making the detailed argument twice, so here, in this introduction, I will only suggest that the *ubiquity* of statements linking fitness or adaptation to such concepts as survival may be taken as *prima facie* evidence of content. Why would scientists have bandied about phrases and propositions about fitness with such regularity unless such talk is useful as they attempt to observe and theorize?[6]

### (2) Logical paradox

Logical contradiction sometimes seems to result from the analysis of events in which chance affects the concrete parameters with which fitness is reckoned. The best-known defense against such paradox is Mills and Beatty's 1979 propensity interpretation of fitness, which seeks to avoid calling (e.g.) an organism struck dead by lightning before reaching reproductive age less fit than its long-lived identical twin. (Mills and Beatty appeal to our intuition by arguing that if organisms are identical overall, then they must also be identical with respect to fitness. This will be recognized as one of the foundational criteria of supervenience.) But I will argue that just as "real" expectations of longevity and reproductive success can fail to be met, so too can propensities be misleading. A more potent means of avoiding paradoxes arising from the agency of chance is to make clear that our understanding of fitness is self-referentially cumulative (i.e., recursive), meaning that we expect to observe fitness at any moment as some function of the fitnesses we have observed previously. If the function which describes such expectations allows for random occurrences, then no logical paradox will emerge from the case of the twins or analogous scenarios. In case a specific expectation is not met, the most general understanding of fitness as *a* function of past performance remains intact; only very *specific* functions must be thrown out. This interpretation is close to that of Mills and Beatty, but it places greater emphasis on the generality achieved by building a recursive understanding of fitness in which the number of iterations can be increased at will to make the impact of the "random" event less marked.

Imagine you heard that on the same day the "unsinkable" *Titanic* left port, another much older, much more fragile ship also set out along the same general route. That the *Titanic* sank while the lesser ship reached its destination safe and sound might be called a fluke if we know of no other details. But if we know every intricate particular of construction and maneuver from the day the respective designs of the two vessels were first conceived up to the moment of the Titanic's sinking, and if we know everything there is to know about the training and psychology of all the crew members involved, *perhaps* the randomness will evaporate. That will not necessarily be the case; it depends on whether one believes (probably contrary to evidence from experiments on the quantum level) that all randomness is merely apparent, resulting from incomplete knowledge of initial conditions and operative forces. But just maybe and in some cases, the ability to convert a circumstance once known only as a staccato progression of sparse data points into a continuum of data with virtually no significant gaps may erase an initial impression of randomness. Arguably some formal ways of characterizing situations are less amenable to such "continuization" than others. If so, then recursive descriptions are certainly on the positive end of the spectrum: they accommodate any increase in the frequency and number of observations. (We will also discuss the possibility of "real" randomness -- that kind which cannot be diminished or removed in the epistemological realm by increasing the knowledge of observers.)

## (3) Interpreting claims about the supervenience and reducibility of fitness

A recursive interpretation of fitness also provides a new approach to questions surrounding the supervenience and reducibility of fitness. If fitness is defined in terms of itself, so to speak, then it cannot be said to emerge *wholly and directly* from concrete circumstances of the last iteration, nor can it be reduced only to such circumstances, despite the fact that there is a necessity relation between them and the fitness value which they helped effect. One might say that recursive definitions blur the distinction between algorithm and argument (in Frege's 1891 sense), so that when we define an X as a function of itself, X = f(X), it is not quite clear what stands inside the parentheses. If we understand supervenience as a multiply realizable dependence between concrete circumstances and a concept, and if we take "microreduction" (cf. Weber 1996) as a means of uniquely associating an instance of that concept with a subset of the set of all circumstances which can realize that concept, then it appears that fitness is supervenient (there is a necessity relation between concrete circumstances and the concept in question), whereas it is not quite clear that the sufficiency criterion of even a microreduction (the presence of a *unique* relationship between an instance of the concept and a subset of realizing circumstances) can be met. This follows from the fact that the recursive algorithm takes *itself* as argument and therefore makes *itself* one of the circumstances of its own instantiation. It may be that from such discussion we can also reappraise the claim that biology is independent of "more basic" physical sciences.

### (4) Arguments over the unit of selection

The "unit of selection" is sometimes understood as the level at which selective forces operate, at other times as the level at which self-reproducing entities exist. Arguably consistent with both understandings is Dawkins' (1976) notion of the selfish gene (based at least in part on the ideas of G.C. Williams, 1966). By this account, the "level" at which the struggle to survive and reproduce plays out is that of the gene, for in the spectrum of levels genes are the only replicators of themselves (no anthropopathic intention is implied). To serve their goal of self-replication, genes make bodies (individuals) to serve as vehicles. Clearly this renders it impossible to view something like DNA as a functional part of an individual organism in any traditional sense. Genetic material does not serve the organism as a whole, but rather the other way around (Hampe and Morgan 1988: esp. 119).

But we can play with a tactic that calls Dawkins' scheme and its competitors into question.[7] Suppose we translate the unit of *selection* or *replication* into the unit of foundational *fitness*, where fitness is conceived as a recursive function and algorithm. In other words, the gene is ultimately the foundational entity which, in combination with other such units, determines the *fitness* of entities at higher levels of organic existence. Through this means we will explore the relationship of recursive fitness to the unit of selection issue at greater length below. For the moment it should be borne in mind that unless genes have *always* existed (which to my knowledge no one has ever argued), then the gene has not always been the unit of replication, ergo it has not always been the unit of selection or fitness. If we want a maximally general concept of selection, replication, and fitness, therefore, we have to find a way of describing the unit associated with these functions so that it can account for the period in which life as we know it -- based in self-replicating DNA, portions of which we call genes -- was still emerging. By any account of how life began (whether mechanistic or not), there was a transition between "stuff" which could not replicate itself and stuff which could (perhaps buffered by stuff which could but did not). For the same reason supporting a recursive understanding of fitness in scenarios involving stochastic processes (q.v. (3) above) -- an inherently unlimited formal ability to increase the number and frequency of observations -- a recursive concept of fitness will be particularly well-suited for an asymptotic approach to the borderline between the non-living and the living, and for accommodating self-replicating units other than genes in case theorists find it convenient to characterize very early lifeforms as being other than genetically based.

### (5) How large a role does (or should) adaptationism play in Darwinian evolution?

A reading of Dennett (1995) makes clear how emotional the debate between the those we can call adaptationists and bauplanists can be. The central question is whether all phenotypes have been maintained or have evolved because of the adaptive advantage which they or their predecessor phenotypes afforded, or whether some phenotypes may in fact be "forced" into existence by broader structural subcontexts of the organisms. (The classic statement of the bauplanist argument is Gould and Lewontin, 1979.) It does not require a novel interpretation of fitness to reason that one can believe in *Baupläne* which "force" certain phenotypes while remaining consistently Darwinian. After all, whether forced or adaptive, all phenotypes have ultimately resulted from the interplay of organisms with their selective environments.

By the bauplanist account, some phenotypes are simply indirectly, even accidentally, related to this interplay. But that is slim comfort for committed adaptationists. A recursive interpretation of fitness cannot solve the basic argument, but it can be used to emphasize that bauplanism and Darwinism are consistent. That ability is not unique to a recursive reading, but we will see that recursion is an elegant means of filling the bill.

### (6) How gradualistic must Darwinian evolution be?

Gould is again in the thick of the debate between the camp which believes that evolution must proceed by gradual, perhaps even imperceptible stages and those who argue that developmental leaps are not just possible, but are also more consistent than gradualism with the fossil record and with what we know about genetic mutation. Once again, recursive fitness cannot resolve the debate, but it functions well for either side.

## 3. The Specific Agenda

### (1) Organization

The brief reflections above suggest the content of our agenda; now to the organization. First a groundwork of concepts necessary to a recursive understanding of fitness is presented. These concepts include multiple realizability (chapter one), recursion itself explicated through a famous computer scientist's reflections (chapter two), and the issues of reduction and supervenience (chapter three). By the end of the fourth chapter, I hope the reader will accept that many currently "hot" questions regarding the reduction of propositions about fitness and the supervenience of fitness are more or less dependent on the concept of fitness one happens to embrace. Reducing a recursively defined fitness value can be taken to mean nothing more than accepting that self-referential definition. Similarly, recursive fitness values are naturally supervenient -- on themselves, again in a self-referential way.

As has already been said, recursive fitness would hold little interest if it did not in some way improve our understanding of evolutionary biology as a whole, which means that there must be something incomplete in the received interpretations of fitness. The second part of the dissertation focuses especially on the role of chance in evaluations of fitness. To that end it is natural to consider Mills and Beatty's famous propensity interpretation (1979). Once again it appears that current goals -- notably construction of the most "chance-proof" conception of fitness and a defense of propositions involving fitness against allegations of circularity -- are served by adopting a recursive understanding of fitness. In so far as stochastic relationships (particularly under a logical relation theory of probability in general) can themselves be developed recursively, a recursive understanding of fitness arguably functions better than a propensity model. As for circularity, a recursive understanding already *is* circular to the extent that it depends on self-reference.

The third part of the dissertation is largely an attempt to reinforce the ground already gained in the first two. The most important rhetorical step of this final section is made in the last chapter, the dissertation's conclusion, in which the notion of recursive fitness is applied to a debate in a field other than evolutionary biology. In this case that field is historical linguistics. Arguably all historical disciplines must construct models with the help of which gaps between hard data points can be filled-in through inference. Moreover, such a model's ontology must include motive forces (such as natural selection) which serve to explain how data points are connected. Consistent with the rest of the dissertation, the primary claim here is that recursive fitness can aid in making sense of a particular debate among historical phonologists. The question is whether the sound changes traditionally united under the title of the Great Vowel Shift constituted a single phenomenon or not. This particular debate lends itself to the purpose of exercising recursive fitness in part because one of its primary participants, historical linguist Roger Lass, frequently appeals to evolutionary biology to make his points. I argue that the debate can be made essentially superfluous if a recursive sense of fitness is adopted.

### (2) Recursive fitness

Before leaving this introduction I would like to offer a very brief exposition of recursive fitness as I conceive it. The reader may find it helpful to bear this sketch in mind while wading through the preliminaries in the next section.

Probably the most important thing to be said about this conception is that it views fitness as being in itself an essentially content-*less* but nonetheless useful concept. The extent to which it has content at all can be described in a surprisingly analytic way: at any given time, the fitness of an organism is some function of its own past fitness (in the same selective environment). Thus recursive fitness plays the role of a ratchet function -- a way of schematizing data in such a way that they display directionality without resorting to a full-blown teleological account which portrays directionality by specifying an end point (cf. Dewey 1922: 262). To say that the way in which we recognize the present fitness level of an entity (such as an individual or one of its phenotypes) is related in some fashion to the way we already recognized the preceding level and will recognize a successive fitness level (assuming the entity in question continues to engage our interest) certainly does not imply a particular endpoint. The basic self-referential form of recursive fitness can therefore be offered quite simply:

$$\text{fitness}_{\text{time}=t} = f(\ \text{fitness}_{\text{time}=t-1}\ )$$

Notice that an initial value is also neither given nor implied in the representation above. That fact may prompt the allegation that recursive fitness thereby fails an important and rather obvious test of correspondence with the methodology of working biologists. After all, researchers do associate definite fitness values to various "levels" of life, including genotypes, phenotypes, individuals, and groups such as taxa. But in fact these fitness values are assigned on the basis of incomplete data sets, or even inferred from already existing values established statistically for other entities, whether on the same or a different level than that occupied by the object whose fitness is under consideration. Thus practical fitness judgments do indeed have a starting point, but it is an arbitrary one. Demanding to know the initial argument of the function given above, however, is analogous to insisting that the "chicken and egg" question be given a definitive, absolute answer.

Representing fitness as a recursive function may disappoint some who expect more from fitness in general than an isolated formalism. Although fitness values are almost universally recognized as multiply realizable, there seems to remain a lingering belief that the fitness demanded by Darwinian evolution must equate to a spectrum of abilities in recognized aspects of the struggle for survival. By this implicit outlook, if nature is really red in tooth and claw, then fitness must have to do with things like sharper teeth and longer claws. Further, some will demand that the conception of fitness employed in evolutionary biology somehow associate itself uniquely with, say, a non-teleological outlook while foreclosing any possibility of association with an opposing view. (The illicit introduction of specific content into what should be pure formalism, and the implicit relation of the formalism under consideration to other formalisms, are problems which are not unique to evolutionary biology. A frequent complaint in hermeneutics, for instance, is that great thinkers' formal systems have been distorted through exemplification and association. Along these lines, Sorrell laments that Hobbes's views on power have been distorted by a long but unnecessary hermeneutic relationship with egoism. Sorrell's dramatic means of countering this connection is to show that Mother Teresa's life's work can be viewed as consistent with Hobbes's views on power! Q.v. Sorrell 1986: 104.) Recursive fitness does not entail comparison in any *particular* modalities at all. Certainly a specific fitness value has to be reckoned on the basis of performance in some aspect of life, but the spectra in which differential performance can occur should not be specified in any way by a formal conception of fitness. Accordingly, such realms of competition are not indicated in any way by the functional description of recursive fitness offered above.

It may be objected that practicing biologists do indeed reckon fitness in terms of performance in categories of endeavor such as longevity and reproduction. Therefore, the objection would go, a wholly content-less conception of fitness would miss the mark in one of the important criteria of evaluation we have already seen, namely correspondence with the practical methods and aims of working scientists. By this account something like Mills and Beatty's propensity interpretation -- which is a propensity *to longevity or reproductive success* -- strikes the right balance between generality and specificity. The motivation behind this objection is praiseworthy, but we will see that a propensity interpretation misses the mark of generality which it has set for itself. This is because a propensity to perform in a certain way can fail to be realized just as an expectation of concrete performance can fail to be justified. We can either live with this type of problem or else devise a wholly formal definition of fitness.

It is also important to note that the basic representation of recursive fitness above does not offer a natural epistemological stopping point, a circumstance in which an entity's fitness is considered to be completely known. This aspect of recursive fitness is consistent with the view that judgments of qualities such as fitness are always subject to revision based on new or newly interpreted data. Specific values of recursive fitness corresponding to actually existing levels of life must indeed be statistically grounded, but below it will be argued that the underlying theory of probability should be something like the so-called logical relation theory (associated with Berenson, Keynes and Carnap) rather than a finite frequency theory (Reichenbach).

Finally, and perhaps most importantly, it should be emphasized that recursive fitness is *a* way of characterizing the ratchet function associated with evolution by natural selection. Since (as we have seen) recursive fitness is content-less and wholly formal, it would be unreasonable to assert that it is the *sole* way of characterizing fitness. But I hope that recursive fitness will emerge as an interesting way of dealing with various issues currently of interest to evolutionary biologists and philosophers of science.

# Chapter Two: A Taxonomy of Multiple Realizability

On the 8th of June, 1855, Charles Darwin penned a letter to the American botanist Asa Gray. "I do not know, whether it has struck you, but it has me," Darwin wrote, "that it would be adviseable for Botanists to give in *whole numbers*, as well as in the lowest fraction, the proportional numbers of the Families." Ever interested in raw facts about his world, the English naturalist wanted to learn how many indigenous plants belonged to various taxa. Thanks to Gray he was aware that about two percent of all native species in a certain area of the United States belonged to the family Umbelliferae, but he wanted to point out to his American colleague that mere knowledge of percentages is inherently unsatisfying for the active observer of nature. "[F]or without one knows the *whole* numbers, one cannot judge how really close the number of the plants of the same family are in two distant countries; but very likely you may think this superfluous" (Burkhardt 1996: 143 - 144).[8]

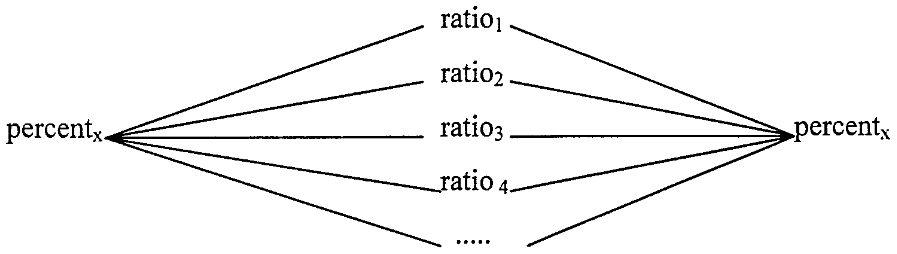
To Darwin the issue was not at all superfluous because he intended to look for numerical regularities existing between plant populations on different sides of the Atlantic. For that purpose, only possession of the raw data would suffice. The text of the letter leaves Darwin's precise intentions vague, but it is clear that he was determined to review an actual count of plant species in various categories such as taxon (in this case family) and domestic status (indigenous or introduced). By having available the raw quantities which, as parts of ratios, constituted percentage values, Darwin apparently felt himself able to perform further calculations as a way of testing hypotheses about intercontinental patterns of species distribution. To repeat, such tests would have been impossible had he known only the percentage numbers or what he calls the "lowest fractions." (It is not hard to imagine questions about plant distributions which might require knowledge of actual numbers. Two naturalists, one living in Middlesex County, Massachusetts, and the other in Surrey, might find it interesting that the Umbelliferae constitute 2 percent of all the native species in their respective areas. But one or both might want to know as well if one area has more Umbelliferae -- meaning the actual number of species -- than the other.)

Gray, incidentally, does not enjoy much "name recognition" these days, but he was perhaps Darwin's most ardent supporter in America against the likes of skeptics such as J. L. R. Agassiz. He was even known as Darwin's "American Bulldog," or in other words as the overseas analog of England's T. H. Huxley (Bowlby 1990: 351 - 2). In the generation after Darwin, however, Gray became almost a counterpoint to a wholly mechanistic reading of Darwinism. His name symbolized the attempt to reconcile evolution with supernatural design (Dewey 1910: 12; Clark 1984: 133). It is possible that to the extent Darwin's own sometimes elusive remarks on the subject can be interpreted as making a place for divine design, Gray was a major catalyst -- for instance in *The Variations of Animals and Plants under Domestication* (1868), which Desmond and Moore suggest "put paid to Gray's divine design as the cause of variation" (1991: 550).

One can read the entire agenda of well-known spokesmen into their every casual remark, but in this case it seems possible that Darwin's implied criticism of Gray is indeed emblematic not so much of a fundamental difference in their respective inclinations (Gray was not uninterested in detail, after all), but of Darwin's own method. He makes a wonderful archetype for the mentality that delights in details, and the closer such facts are to "raw" observations, the better. Of course Darwin had a yen for making inductive generalizations based on such facts (else we might not have the *Origin*), but much of his taxonomic work is solidly based on relatively unembellished observation. In the letter quoted above, the implicit appraisal of Gray (as someone to whom a question of basic fact -- How many Umbelliferae? -- may seem superfluous in comparison with a query of slightly higher order -- What percentage of native plants do the Umbelliferae constitute?) suggests that Darwin shunned the cobbling together of layered generalizations, in which the foundation of raw fact could become remote or even be forgotten. That same concern will be at the heart of this chapter as we explore various relationships which can exist between data and generalization.

## 1. Definitions and terminology

On the face of it, the letter to Gray is simply another manifestation of Darwin's well-known affinity for detail. But considered more abstractly, Darwin's suspicion of percentages seems to stem from a ubiquitous side-effect of measurement: sometimes the analysis and manipulation of raw data produce what we might call "one-way" quantitative results. Knowing that two percent of the indigenous plants in a certain region belong to a particular family is interesting in its own right, but it may not be sufficient for all purposes. As just noted, if we wished to compare the numbers of Umbelliferae in Middlesex and Surrey, we would need to have a census of each family's members rather than just the percentage value of each family as compared with all other native plants in a given area. In general, a percentage represents a proportion which could have been yielded by any number of numerators and denominators, just as any number of triangles might be congruent to one another; in particular, a value of two percent in the case concerning the Umbelliferae compared with other native plants corresponds to an infinite number of ratios -- one member of the family compared with fifty indigenous species in all, or 2 versus 100, or 3 out of 150, and so on. In that sense, values expressed *per centum* are realizable in many ways. A single, definite percentage value corresponds to a given ratio, but not vice versa. Or put another way, the relationship between the set of percentage values and the set of ratios is not one-to-one. A picture is not especially enlightening in the case of such a simple relation, but a comparison of graphic representations may help maintain clarity as refinements are made later in the chapter. What we have is a sort of a "fanning-out" relationship between elements of the set of all percentage values and elements of the set of all ratios.



As the diagram indicates, there are an infinite number of ratios corresponding to any percentage value. Moreover, the "direction" of relationship can be seen as moving in either direction; a given percentage value can be "reduced" to any of a number of ratios, and all the members of a certain set of ratios can be "reduced" to a given percentage.

In modern philosophical parlance the general phenomenon represented here is sometimes called "multiple-realizability" or even "supervenience," although not all philosophers of science are comfortable treating the two terms as synonymous (witness Weber's evaluation of Sober: "Sober uses 'supervenient' pretty much synonymously with 'multiply realizable,' which is not quite correct, since two properties that are correlated in a one-one manner also satisfy the supervenience definition" -- Weber 1996: 417). Let us defer discussion of supervenience for the moment and concentrate on multiple realizability. The meaning of this term can probably be made as clear through heuristic means as through a more didactic presentation, but we should proceed carefully nonetheless. Here, then, is a first stab at a general definition:

(1.a) A phenomenon is multiply realizable if more than one cause suffices for its instantiation. Or to describe the same definition a bit differently, a circumstance is multiply realizable if none of its (multiple) sufficient causes is necessary.

## (1) A rough taxonomy of causes

Since the concept of a cause figures importantly in this definition, it will be worth our while to consider briefly what the concept can mean. A good straw-man outline of causal types is Aristotle's taxonomy.

## (a) Material cause

By "material cause" Aristotle meant the specific substance which constitutes a thing. "In one sense, 'a cause' means (1) that from which, as a constituent, something is generated; for example, the bronze is a cause of the statue, and the silver, of the cup, and the genera of these [are also causes]" (194b24-26, Apostle (tr) 1980: 29; Apostle's square brackets).

Can fitness function as material cause or can a given fitness value be multiply realized by various material causes? Fitness differentials are sometimes attributed to variance in the stuff of which structures are made. Presumably an ankylosaur of the Cretaceous Period could have had the same form if it were sheathed in normal hide rather than being "encased in armor" (Futuyma 1986: 336). But the material difference in its relatively impenetrable skin as compared with the softer surface of some of its peer species presumably gave it an advantage in survival and reproduction. This is not to say tougher hide always yields these advantages, but in this particular case it is conceivable that we would consider the armor to be fitness-enhancing.

Our concept of material cause and of fitness as such an agent becomes more complicated if we consider whether more than one specification of the substance in question is possible in a given case. If there is more than one possibility, then we must ask whether any single description is privileged, that is, to be desired over all others. Apparently there is no single description which is always and uniquely appropriate. This will become clearer if we take an example. Suppose we identify bronze as the material cause of a certain statue. Aristotle's text (quoted above) makes it clear that we can consider *this* (particular) bronze as well as bronze generally speaking to be the matter of which the statue is made. But presumably we could specify the material component even more broadly by stipulating "metal" instead of "bronze," for instance, or more specifically by offering the exact chemical composition of the particular statue's bronze. This last way of describing the material involved need not be analytic, since some degree of variation is possible: what we call bronze is *normally* a product of copper and tin in a certain ratio, but there is some latitude in that mix. *Webster's* goes so far as to say that tin may be entirely absent from bronze. Moreover, whether tin is present or not there are "sometimes other elements" in the alloy (*Webster's New Collegiate Dictionary* 1974: 141).

The possibility that the material substrate of a given object may be variously identified is significant for the concept of fitness as material cause. We might encounter a statement of the form "X lives longer or reproduces more successfully than Y because X is fitter than Y." In turn, we could entertain the theory that X is fitter than Y because of the material of which X or some part of X is made, just as the ankylosaur's armored body may have made that dinosaur fitter than some of its softer-skinned peers within the specific environment which all occupied. In this case there is no apparent reason to fear that we may have misidentified the matter that is tied-up with fitness. It makes sense that having body armor could enhance fitness, extra weight and the challenge of temperature control notwithstanding. But we can never be certain. Suppose we question a given material's contribution to its possessor's fitness. Is a shark better off with a skeleton made of cartilage rather than one made of bone? On the one hand, "the abundance of cartilage in sharks may explain why these fish are not prone to cancer" (Lane and Comac 1992: 37), but on the other hand, "[b]ecause a shark does not have a rigid skeleton or a rib cage to protect its internal organs, it can be killed by a porpoise -- one of its normal food sources -- in a one-on-one fight. The porpoise can literally butt the shark to death since there is no rib cage to protect the shark's vital organs" (ibid.: 10).[9]

Where each of the rival materials affords benefits as well as drawbacks, questions about fitness value may remain moot. But it is tempting to believe that a definitive resolution could be found if we were able to identify the relevant material causes in a different way. In our present example, if it is true that sharks almost never suffer from cancer, is it cartilage *per se* which affords this benefit or is the telling factor merely associated with cartilage in some fashion? (To my knowledge the precise cause of the low incidence of cancer among sharks has not been identified.)

The import of these speculations for our present purposes can be summarized in the form of a question: Assuming fitness can function as a material cause in some explanations or predictions, can it be equated with any *particular* material? Indeed fitness seems sometimes to be identified with an Aristotelian material cause, but such a connection is problematic even in specific, limited contexts where the issue of identity versus supervenience apparently can be set aside. The problem arises in specifying the material which allegedly confers advantage in reproduction and survival. Although we may be confident that such benefits are in some manner associated with a material feature described at a certain level of specificity (e.g., as "bronze" is more specific than "metal" but less specific than a particular ratio of copper, tin and other alloys) in many cases -- or arguably even in all instances -- we must always be left wondering whether the fitness advantage is not more closely associated with some other level. Thus we can say that the material cause *underdetermines* fitness in somewhat the same way that data are said to underdetermine scientific theories in general. A material cause can function sufficiently without being necessary, in perfect conformity to (1.a) above.

### (b) Formal cause

Aristotle's introduction of *formal* cause is evocative of an organism's structure at any number of levels. "In another [sense, cause]...means (2) the form or the pattern, this being the formula of the essence, and also the genera of this..." (194b27-30, Apostle (tr) 1980: 29). By "the form or the pattern" (το ειδος και το παραδειγμα) we might understand something like the cellular organization of an organism which allows it to live in one environment but not in another. Concluding a discussion of the tropical alga *Caulerpa*, the largest known single-celled organism, Jacobs remarks:

> If Caulerpa is this prominent as a large, highly differentiated, multinucleated single cell, what are the ultimate lengths to which this structure could be carried? I can see nothing that would preclude an even larger unicellular organism so long as it lives in the sea. There the buoyant water substitutes for the internal support provided to land plants by their cell walls (Jacobs 1994: 105).

Or a structure associated with fitness might be at a higher level:

> and flying squirrels have their limbs and even the base of the tail united by a broad expanse of skin, which serves as a parachute and allows them to glide through the air to an astonishing distance from tree to tree (Darwin 1859: 180).

It is interesting to speculate what constitutes the formal essence of the organisms described in these passages. Caulerpa demonstrates that whereas size may be associated with formal cause among some or even all land plants, size need not be as critical a formal parameter in all environments. Similarly, Darwin took the form of a flying squirrel to demonstrate how "the accumlated effects of this process of natural selection" (ibid.: 181) can undermine a philosophy of natural kinds based on immutable *formal* essences (cf. Mayr 1984: 531 - 534 and Futuyma 1983, but cf. Lennox 1987 for a contrasting reading of Aristotle in this context). What happens in the Darwinian concept of speciation is that form loses its monolithic character. Descartes asked of a piece of wax, once hard and cold and of distinct shape, but now runny and warm and amorphous: "Is it still the same wax despite these changes?" He answered his own question with certainty: "We must admit that it is; no one can deny it, no one can think otherwise" (1961: 67). But Darwin's m.o. can be taken as denying the essentiality of formal cause as an essence of particular organisms. Thus the kind of cause which under some conceptions (e.g., traditional readings of Aristotle) would be necessarily as well as sufficiently associated with an organism turn out to figure only in sufficiency relationships.

### (c) Efficient cause

Aristotle's concept of an *efficient* cause is given by the following passage.

> "In another [sense, cause] means (3) that from which change or coming to rest first begins; for example, the adviser is a cause, and the father is the cause of the baby, and, in general, that which acts is a cause of that which is acted upon, and that which brings about a change is a cause of that which is being changed" (196b30-32; Apostle (tr) 1980: 29-30).

The hallmark of this explication is that it is extremely broad. In other words, it is tough to think of a kind of cause which does *not* in some way to conform to Aristotle's description of efficient causality in this passage. Once again, this kind of cause can be sufficient but is certainly not necessary for the roles offered as examples.

### (d) Final cause

It is tempting to conflate the *final* and *formal* causes of, say, a statue -- to claim, for instance, that the statue's finished state is the "final cause" to which it was always tending (cf. Lass 1976: 54). But Aristotle seems to have had something else in mind. "Finally, [['a cause']] means (4) the end, and this is the final cause [that for the sake of which]; for example, walking is for the sake of health. Why does he walk? We answer, 'In order to be healthy'; and having spoken thus, we think that we have given the cause" (194b33-34; Apostle (tr) 1980: 30; his single square brackets, my double ones). Surely we can identify final causes of the statue in the sense of things "for the sake of which" (to use Apostle's phrase) it exists: human aesthetic sensibility if the statue graces a museum gallery, the desire to inspire the People to further revolutionary glories if it looms over Tiananmen Square, the evocation of envy if it confronts visitors to a conspicuous consumer's drawing room, or simply the ability to hold down a stack of paper when an office window is open. Fitness seems to fill this role as well, and in fairly obvious fashion. But once again we see that the final cause need not be necessary despite its sufficiency for the various roles.

### (e) Summary of the causal roles: necessary for specific cases, sufficient for general ones

The discussion of the four causes above makes it clear that all share at least one characteristic: they can be necessary for a *particular* entity's coming into being and for its continued existence. A certain object such as a statue cannot begin to exist unless there is material like bronze out of which it can be formed, and the bronze remains a *causa essendi* as long as the statue exists. Similarly, a *specific* statue would not be what it is without its form. We cannot rightly say that in the absence of this form the statue would be something else -- a lump of bronze, say, or a different statue or a number of teapots. Rather, that *particular* statue would not exist, period. The same would be true had the sculptor, whom we can identify as an efficient cause, not existed. Without such final causes, the statue would likewise never have come into existence.

Thus the four causes, taken singly or as a group, can justifiably be called "necessary" (hence the title of this chapter) in so far as they are that without which the *particular* thing in question would not exist. But they are sufficient rather than necessary for the particular *kind* of thing (e.g., statue). It is important to note that the relationship of necessity is one-way. In order to exist, the statue requires the material in which it is manifested, but not vice versa; the material can exist in another form. Similarly, the statue must have the form which it bears in order to be that particular statue, but the form could have been realized through holographic means. Further, the sculptor might just as well have gone to work designing auto bodies for Ford, and a nice watercolor would conceivably have gratified someone's aesthetic sensibility as well as the sculpture did. In short, we see that the existence of a particular thing T presupposes the existence of four causes:

(F.1) $$T \Rightarrow \{M, Fo, E, Fi\}$$

but not vice versa:

(F.2) $$\sim [\, T \Leftarrow \{M, Fo, E, Fi\}].$$

Admittedly this conclusion may be seen as problematic, particularly in the case of the formal cause. One might argue that if we specify the form with enough precision, only that particular statue will fill the bill. In the same way, an aesthetic final cause might be described so thoroughly that there is only one object which could reasonably be thought to fulfill it. If a museum director phoned the sculptor and said, "We're going to do something in the East Wing this coming spring, and I wondered if you had anything with atomic number 79," then presumably the sculptor could only offer a piece made of gold. But such attempts to establish a biconditional relationship between a causal type and its effect seem too problematic to serve as the foundation of a deterministic theory of causality in general.

Or we can demonstrate the non-deterministic character of the four causes in a different way. Each of the four causal types introduced above can be understood as a set, while specific instances of each can be grouped into subsets of subsets. For example, we might conceive of the set of all material causes, whose subsets could be specified as animal, mineral, and vegetable. These three subsets are amenable to further division: the animal heading might subsume the phyla which most taxonomists currently group under the kingdoms *Animalia* and *Protista*.

Aristotle's text makes it clear that a given phenomenon can be linked to more than one cause. For instance, we saw above that in the case of bronze or silver as the respective material causes of a statue or cup, we can identify not just a particular instance of the metal but also the metal in general. In other words, one may distinguish between the specific material which the statue or cup comprises and the genus of that material: *this* bronze or silver versus bronze or silver in general (q.v. Apostle 1980: 209, n. 3).

There may be some question as to the causal category corresponding to a given phenomenon. For instance, under the heading of material cause above we briefly considered the possible fitness advantage accruing from armor-like skin. In the context of Futuyma's mention of this kind of hide (1980: 336), it seems most natural to treat such skin as a material cause to the extent that it is seen as agent of a particular effect. But it would be possible to conceive of hide as a formal cause, as well, in so far as we could reduce the significance of the material to a matter of molecular form.

In summary, it seems that a given fitness value may be multiply realized by any of the four types of cause in Aristotle's basic taxonomy, since there is no biconditional relationship associated with any of the four causal categories and their particular or general effects. In the discussion below we may from time to time need to comment on the particular type of realizing substrate which is under consideration.

### (2) Eyewitness codicils and the question of inevitable multiple realizability

We should note carefully what definition (1.a) does *not* stipulate: It does not entail that we must be ignorant of what caused a specific multiply realizable effect. On the contrary, it is perfectly conceivable that an observer can identify the particular cause of a multiply realizable phenomenon in some cases. Any number of factors might be responsible for the sensation I currently feel on my foot and lower leg, but peeking under the desk I see that a gregarious neighborhood cat (which tends to slink through my window everyday about this time) is playing with my shoelace. The sensation is multiply realizable (conceivably a similarly playful puppy or a blustery wind or a neurological disturbance could have been responsible for the same intentional state), but in this case I know its cause. If I had not looked under the desk while the cat was still present, however, I might never have discovered that it was the actual agent of that peculiar sensation. Similarly, Darwin *happened* to know that in the region of the United States which Gray had studied, the actual number of native Umbelliferae compared with that of other indigenous species was 36 versus 1798. In turn, the raw numbers could be reduced to about 1 in 50, or two percent. We have already seen that had Darwin not known the raw numbers, he could never have inferred them from the percentage value. Hence his complaint to Gray, whom he feared had not seriously considered the matter (as evidenced by the suggestion that Gray might find the issue to be "superfluous"). But the mere fact that a percentage value does not correspond to a specific ratio in general does not mean that we might not possess sense data or some other kind of special knowledge which would allow us to link the indefinite to the definite, to etch a unique relationship against a many-to-one backdrop.

Let us call information which we *happen* to have about a multiply realizable phenomenon such as a percentage value or the sensation caused by the cat batting my shoe lace around with its paw an "eyewitness codicil." The significance of this kind of information will be made clear momentarily. For the time being let us agree as a matter of convention that the elements in a subset of all multiply realizable phenomena are also what we might call "one-way" in the following sense:

(1.b) One-way phenomena are multiply realizable occurrences which cannot be reduced to a *single* equivalent expression in another "category."

Thus, for instance, if Darwin had known only that 2 percent of all native species in a given area belonged to the Umbelliferae -- that is, if he did not know the raw numbers which yielded that percentage value -- then the figure "2 percent" (representing the category of percentage values) would be one-way with respect to the category of all ratios.

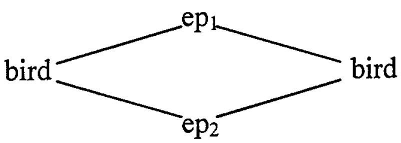
A question immediately presents itself: Is there such a thing as an inevitably one-way phenomenon, or is it always possible that an eyewitness codicil can happen to apply? Perhaps, as the example of the cat rubbing itself against my leg suggests, ability or inability to identify the specific cause of the sensation *always* has to do with what I *happen* to know or not know. If I should chance to see the cat at the same time I experience the sensation, then I can explain the feeling by reference to that specific cat (and if not, then not). Presumably such an eyewitness codicil might apply in many cases of multiple realizability, but there seem to be two major classes of circumstances in which we cannot *happen* to observe causal connections. First, there is the possibility that in cases such as that demonstrated by the two-hole experiment (Feynman 1965: 127 - 148; Gribbin 1984: 164 - 176), some happenings are inherently unpredictable. Even if we know everything there is to know about initial conditions, we still cannot tell which of a range of possibilities will prove to be the case. And if we cannot *predict*, then perhaps we also cannot *reduce* from an occurrence all the way back through each intermediate stage to a set of initial circumstances (i.e., causes). But although quantum phenomena may be relevant to living organisms, albeit in a way many times removed, the locus of our discussion of fitness seems to be at a different "level" (an admittedly vague term which I hope the reader's intuition will suffice to comprehend). So we will leave this realm of events out of consideration for the time being.

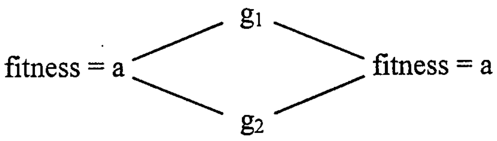
Secondly, there seems to be a kind of temporal boundary along the path between the set of possible realizing phenomena and the multiply realizable occurrence corresponding to that set. This boundary entails that even where there is no unpredictability of the kind demonstrated by the two-hole experiment, the observer still may be unable to reduce *after the fact* or *in general*. In other words, if we observe a certain phenomenon, we may be able to say that it did occur in only one way (*that* way -- the way we observed). But after the fact, when we observe not a dynamic process but rather a static phenomenon which we think must have resulted from some chain of past events, there is arguably always an element of uncertainty. (Usually that feeling results from the cat batting my shoelace around. But since I didn't trouble to look, I can't say for sure that it was the cat, even though she's now sleeping a few inches away from my feet.) Moreover, if we cannot link a given event with certainty to a causal predecessor, then it seems we cannot link such events *in general* to definite causes. (In some cases it might be more appropriate to substitute a phrase such as "predecessor state" for "cause" to express the relationship between a multiply realizable phenomenon and the "basis" of circumstances upon which it can be realized. More will be said on this subject below.) Let us leave the basic question (Are any phenomena inherently and irretrievably one-way?) unanswered for the time being, remarking only that many if not all multiply realizable phenomena can be reduced to specific causes whenever an eyewitness codicil applies, but not otherwise.

### (3) Conjunctive and disjunctive multiple realizabilities

It should also be remarked that the definitions of multiply realizable and one-way phenomena  presented in (1.a) and (1.b) above apply to conjunctions as well as disjunctions. That multiple realizability has to do with disjunction is clear. We conceive of a range of possible causes -- or we can use the more general term "predecessor states" to encompass the relationship between entities such as two percent and {0.5/25, 1/50, 2/100, ...} -- any one of which might be linked to a given result. The "fan" diagram above does not tell us whether the individual ratios are singly or collectively linked to the percentage value. Of course in this case we know that the ratios are in some sense the same, so that any one of them could take the place of the percentage values in the diagram. So let us imagine another case, one in which the multiple realizability in question is not quite so many-sided.

Suppose biologists have discovered that two groups of organisms (which need not be well-recognized taxa) have the same fitness value. Or say that paleontologists are divided into two camps, one of which holds that birds are most closely related to dinosaurs while the other side of the debate envisions a different evolutionary path. In these cases we see two possible linkages:





Incidentally, it should be clear that these examples are based on a broad conception of multiple realizability -- perhaps broader than some would think wise. If one takes multiple realizability to be a synonym for supervenience (as Weber 1996: 417 suggests Sober does), then it will not do to say that birds as a taxon are multiply realizable because more than one path of their evolutionary development is conceivable. We would not, after all, wish to say that birds as a group *supervene on* the different hypothetical courses of development. But that should not stop us from treating multiple realizability in a broad way here, since the point is to discover the various forms which can be taken by multiple realizability at its most abstract. We can constrict the definition appropriately if it turns out that a narrower understanding jibes better with concepts such as supervenience.

Unlike the case of percentage values, there is no implication that the realized values (in the middle) of the two diagrams above can be linked together apart from their relationship to the end points. In other words, there is no sense that the evolutionary paths x and y of birds are in any way equivalent, nor are the groups of organisms $g_1$ and $g_2$ the same, whereas the uncountably many ratios corresponding to two percent are themselves somehow similar if not, loosely speaking, equal. The bird example is what could be called an either-or case: Either x is linked with birds or else y is linked with birds, but not both. Of course that conclusion is corrigible, but we can imagine that at any given time the theories represented by x and y would be thought of as mutually exclusive (that is, *either* the evolutionary path of birds descends most directly from dinosaurs *or* it descends most directly from some other group of organisms, *but not both*). Let us call this kind of multiple realizability *true disjunctive*.

Formally, the example of multiply realizable fitness in the diagram above is not quite truly disjunctive. There is, however, a similarity between this case and the one involving the evolutionary development of birds. In both examples, the existence of either alternative suffices to establish the thing represented diagramatically as an end point. But because we can conceive of the possibility that two distinct organisms (or groups of organisms) share the same fitness value, it is clear that in the case of multiply realizable fitness values, *both* linkages ($g_1$ *and* $g_2$ linked to fitness value a) can exist *simultaneously*. Thus there is a certain sense in which the relationship is conjunctive, although the conjunction is "buried" in a disjunction (one alternative *OR* the other alternative, *OR* the first alternative *and* the second alternative). How do we handle such a situation in which multiple propositions involving a single fitness value can all be true (e.g., "$g_1$ has fitness value a" *and* "$g_2$ has fitness value a") but are independent of one another? For clarity's sake let us dub these sorts of instances *simultaneous disjunctive* in order to distinguish them from *true disjunctive* cases on the one hand and from situations in which only some *conjunction* of circumstances is sufficient to cause a multiply realizable phenomenon on the other hand. Conjunctive multiple realizability remains to be illustrated.

Conjunctive relationships are an old worry in philosophical analysis, one for which there seems to be no clear-cut solution. Such conjunctions amount almost to the problem of *defining* something: What conditions must be fulfilled in order that a thing be what it is? When Socrates asked for a definition of virtue, for instance, Meno responded with a *list* of attributes, that is, with a *conjunction* of components -- justice, temperance, piety, etc. Such an answer is unsatisfying, since clearly a composite thing is not equivalent to any one of its parts. Thus Socrates' lament: "...whereas I asked you to give me an account of virtue as a whole, far from telling me what it is itself you say that every action is virtue which exhibits a part of virtue, as if you had already told me what the whole is, so that I should recognize it even if you chop it up into bits" (79b; Hamilton and Cairns 1961: 362).

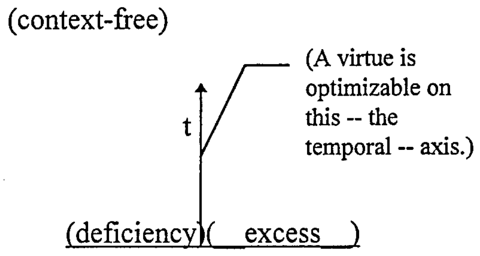
(Appealing to examples from the theory of ethics seems natural in the present discussion of multiple realizability. One of our goals, after all, is to address the alleged supervenience of fitness on concrete physical circumstances. Kim, for one, begins his treatments of supervenience by remarking that the term itself seems to have emerged from the moral theories of G. E. Moore and R. M. Hare (Kim 1978: 149; 1984: 154 - 5; 1990: 3).)

Meno may have offered an inadequate definition of virtue in the abstract, but perhaps he was on the right track to the extent that he envisioned the general case -- let us call it Virtue (with a capital V) -- as being a conglomerate made up of parts -- virtues (with small vs). While Virtue may be some unique combination of parts, perhaps there are phenomena which can be realized by more than one combination of elements in a basis set, but in such a way that none of the units in the combination would be sufficient to effect the multiply realizable phenomenon. Such a case would clearly be distinct from true disjunctive multiple realizability, in which one and only one of a set of basis units suffices to realize a given phenomenon, and from simultaneous disjunctive multiple realizability, in which more than one unit can simultaneously *but independently* suffice to establish the phenomenon in question. In other words, there seems to be a legitimate understanding of multiple realizability which does not conform to the pattern in which more than one cause is sufficient but none is necessary to realize a given phenomenon. This reading, which we can call *conjunctive* multiple realizability, takes some *combination* of independently existing predecessor circumstances as the basis of the phenomenon in question. Each of these "sub-causes" is necessary but none alone is sufficient. Intuitively, that aspect of the pattern seems to describe the sufficiency relationship between phenotypes and fitness as well: no single phenotype, nor any combination of them, is sufficient to make an organism well adapted to all conceivable environments.

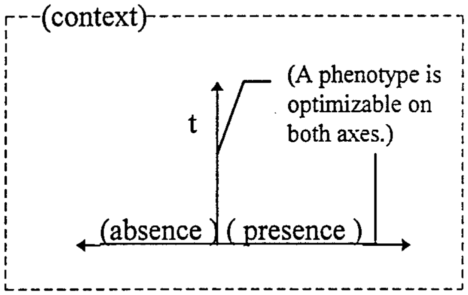
But what about the necessity relationship? Socrates' discussion with Meno makes it clear that while no combination of specific individual virtues is equivalent with virtue in the abstract, all of the particular virtues will in some sense be present (if only *in potentia*) whenever Virtue is present. This is apparently not true of the relationship existing between individual "fit phenotypes" and abstract fitness, even in the same environment. Long legs may be adaptive for some organisms in a given environment but not for other organisms in the same context. Nonetheless, the specific fitness value of a given organism at a given moment in time is realized by a conjunction of phenotypes, none of which is sufficient. (Naturally we assume a given environmental context as well, else the very notion of fitness is meaningless. That points out another difference between the cases of Fitness and Virtue, since the latter is apparently context-independent by the non-relativistic reading which thinkers such as Plato and Aristotle seem to invite.)

It is tempting to speculate that each environment uniquely determines a sort of "Super Organism" -- one which has, if not the best of everything when phenotypes are considered individually, then the best combination of possible phenotypes. Of course such an organism need not actually exist for it to be considered a possibility. If only in theory, this situation might be thought to resemble the relationship of virtues to Virtue with respect to necessity and not just sufficiency. We can assume that whatever Virtue is (and we know the definition remained problematic for Socrates and Plato), it apparently bears some relationship to individuals virtues in which each of these units is necessary but none is sufficient. Similarly, while no single Super Phenotype is *sufficient* for the presence of Super Fitness, so too each of the individual Super Phenotypes would be *necessary* to Super Fitness. Two considerations make such a comparison problematic. First, there is the design problem. How could we design (before the fact) or recognize (in hindsight) the Super Organism? Might it not be the case that numerous designs all approximate the optimum, with performance differences so slight as to make all the options indistinguishable? There appears to be no way of determining whether this is a corrigible problem or not -- that is, whether it constitutes a practical difficulty rather than a metaphysical impossibility -- so let us ignore it for the moment.

The second problem with comparing the relationship between fitness and phenotypes (or any units of the organism) with that between Virtue and virtues is that, in the conception under which thinkers such as Plato and Aristotle labored, there was a sort of natural limit placed on the optimality of virtue (hence the long discussions of moderation, for instance in the *Nicomachean Ethics*, II.2 - 9). The doctrine entails that "too much courage" is an oxymoronic phrase, since "too much of any virtue" is no virtue at all. Courage as a mean is bounded by cowardice and foolhardiness; once those boundaries have been passed, the behavior in question is alloyed and is no longer courageous. There is thus only one "axis of optimality," a temporal one. To be always X (where any virtue can be substituted for X) is the best one can do in the singular, and to be always X, Y, ... etc. (where the variables exhaust all virtues) is the pinnacle in the combined context. This is apparently not true of phenotypes. A quality such as acute vision is indeed optimizable along the temporal axis (conceivably a human being whose power of visual perception were constantly operative would be better adapted to most environments than one who had in effect no vision while sleeping, *ceteris paribus*), but there is virtually no way to limit a property such as vision on what might be called the "quantitative axis" in the way that the Greek doctrine of the Golden Mean meliorated the concept of individual virtues. A metaphor employing Cartesian coordinates can be carried too far, but it seems that we can get some mileage from thinking about the relationship between the multiple realizability of individual virtues and the "fitnesses" of phenotypes in something like the following schematic fashion:

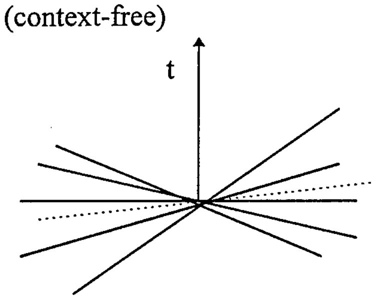


a single virtue

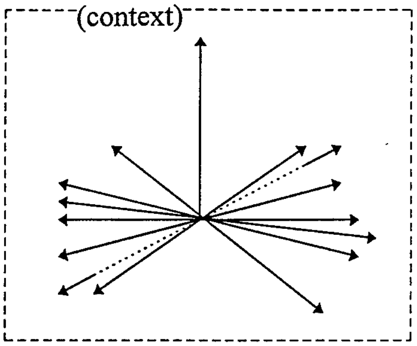


a single phenotype

One can extrapolate these diagrams into others representing Virtue and Fitness (with a capital V and a capital F, notice) using a broken line to represent the fact that it is not entirely clear how many virtues or phenotypes are represented.



Virtue



Fitness

The point of these graphic forays is to emphasize the background of a notational convention which may prove helpful. Let us agree to call multiply realizable phenomena whose constituents themselves (and not just their duration) are optimizable "twice-dimensional," while those phenomena whose units are analytically bounded with respect to their quiddity (such as Greek virtues) will be dubbed "one-dimensional" multiply realizable phenomena. Courage, for instance, is optimizable along the axis of duration (Socrates demonstrated courage before and after being sentenced to death) while an eye can be optimized along what we can call the "axis of quiddity" as well (*ceteris paribus*, that organ of sight is best which sees not just always, even when its possessor sleeps, but also when it is more sensitive to a greater spectrum of radiation than competitor eyes).

As indicated earlier, an organism's fitness in a given environment seems to exemplify a category of phenomena which we might describe as being conjunctively and twice-dimensionally multiply realizable. This means that a given, numerically expressed fitness value can be realized only by a *combination* of units in a basis set (conjunctivity), that at least some of those units can be conceived as optimizable along *both* the quiddity and duration axes (twice-dimensionality), and that *various* conjunctions of differently optimized units can yield that value (multiple realizability). For the sake of convenience, we may choose to speak of fitness as though it were one-dimensional or as if it were disjunctively multiply realizable. Examples are not difficult to find. Recall that in the case of true disjunctive multiple realizability, we have a given state of affairs S, a set of predecessor states (e.g., causes) C: {c1, c2, ... }, and a number of possible pairs S-$c_i$, S-$c_n$, etc., *one* (and only one) of which is an accurate description of a given circumstance. We might hypothesize that a given giraffe was longer-lived and produced more offspring than its peers because it had a longer neck or because it had a longer tail. While both phenotypes may have contributed to the individual's success, it is at least conceivable that only one phenotype was responsible. For instance, the pair{successful-long neck} may reflect a truth about nature while the pair {successful-long tail} does not. But both of these examples assume without expressing the fact that the pivotal phenotypes must exist *in combination* with others. If we manipulate the same example a bit, we can produce a simultaneously disjunctive case as well. Perhaps giraffes are either big and slow or small and fleet. Both combinations may yield the same longevity and success in producing offspring, so that the triplets {big, slow, successful} and {small, fleet, successful} describe a state of affairs actually found in giraffe populations. Once again, however, the triplet does not make sense unless we assume that there are other phenotypes associated with those expressed in the triplet. Thus convenient expressions do not really diverge from the pattern of conjunctive multiple realizability.

Nor do they escape part of Socrates' gripe with Meno: In general, *all* the possible combinations exist (as in the case of simultaneous disjunction), but no combination is sufficient to realize the quality in question; in a particular instance, a given combination is sufficient but not necessary. The explanatory challenge posed by conjunctive multiple realizability is to formulate an umbrella concept which encompasses the essence of all of the generally necessary but insufficient linkages. A longer-necked, faster, keener-sighted, smarter giraffe is fitter than its less endowed peers, but fitness *per se* is not long-neckedness, foot speed, keen-sightedness nor intelligence, nor can we claim that fitness is a combination of any or all until we are certain that the range of possible linkages which we have in mind is exhaustive and until we have a data base linking each combination with a track record of longevity and reproductive success in every conceivable environment. Compiling such an exhaustive data base of linkages is of course impossible; inference will have to do yeoman's work in any body of practical claims in evolutionary biology. But there is a formal problem in addition to this practical one. By describing fitness as conjunctively and twice-dimensionally multiply realizable we have not thereby overcome what I will call "Meno's problem": What is the exact relation of "microfitnesses" to "Fitness"? (This entire dissertation could be viewed as a defense of two theses: first, that the problem is insoluble; second, that the problem can be circumvented for at least some purposes by defining fitness recursively.)

## (4) The theory-ladenness of multiple realizability

In the debate over the meaning of fitness as in science generally, it is possible to misconstrue the range and content of the set of possible causes corresponding to a multiply realizable phenomenon. Sometimes this misapprehension results not from a simple failure to take account of something which available concepts and tools of investigation *could* have made visible, but rather from a more complex, theory-driven inability to look at the right part of nature and to look in the right way. Ambrose (1996) describes the ubiquity of malaria not just on the frontier but in early American life in general ("Jefferson had it") and the foibles of contemporary scientists as they sought a cure for two all-too-common diseases, malaria and yellow fever, which they called the "ague" or "bilious fevers." Part of the problem of investigation was that the extant general theories of disease transmission drove the process of hypothesizing and testing in the wrong direction. Lacking the insight that insects could be vectors of disease, there seemed to be two possible means of transmission. First, having observed that well people sometimes became sick after entering the presence of ague victims, direct bodily contact was thought by some to be responsible for transmission of the disease. The other potential vehicle -- "bad air" -- explained why sometimes people seemed to get sick when they were merely in the vicinity of ague sufferers without having touched them in any manner.

> Dr. Rush's opinion [of the vehicle of transmission] was bad air, rising from the swamps. He was on the edge of seeing the mosquito as the culprit, but never quite got there. The good doctor was helping one of his University of Pennsylvania graduate students in his research. S. Ffirth was doing his thesis on causation of malaria. The general view was that epidemics resulted from 'contagium,' meaning the fevers were transmitted by direct contact between people. Ffirth wanted to see whether this was true...[He] inhaled vapors from black vomit taken from malaria or yellow-fever patients. He injected the vomit into the stomachs and veins of cats and dogs, and into his own body. Neither the cats and dogs nor Ffirth got malaria. He completed his research in June 1804, and reported his conclusion: the 'autumnal disease' (yet another name for malaria) was not contagious" (Ambrose 1996: 113).

Doubtless Ffirth's scientific odyssey contains many morals for graduate students, not the least of which is that rigorous (even extreme) methods do not always yield the truth, even when they are carried out in a way consistent with a sound research strategy. It is not the case that Ffirth violated every modern precept of scientific method. On the contrary, much of what he did seems wholly proper. For instance, his methods seem to conform to Popper's notion that scientific (empirically meaningful) propositions must be falsifiable. Indeed, falsification makes a reasonably accurate one-word description of Ffirth's scientific agenda. But the graduate student reached a false conclusion -- despite an arguably sound scientific method -- because of the range of causal hypotheses which he considered and tested. Vapors of black vomit (which certainly should fill the "bad air" criterion, even for the olfactorily challenged) did not cause malaria, nor did the most intimate of bodily contacts (realized by injecting bodily fluids from sick organisms into well ones). That malaria might be blood-borne, and that its normal vector might be the mosquito, were propositions which Ffirth could not test because he had not conceived them as possible explanations of malaria's spread.

Again a comparison with moral issues may help drive home the point. Daily we see evidence that certain moral stances -- represented by such phenomena as peace treaties, legislation, judicial decisions -- are treated as *revelations* of possibilities never before or only dimly glimpsed. The surprise comes because in either realm, science or ethics, the conceptual locus may be too constrained to allow certain theories room to emerge and grow. But even the strongest conceptual walls may collapse if enough time is allowed *and if theoretical commitments change*, as is made clear by the Vatican's 1984 "pardon" of Galileo or Spain's 1992 announcement that the expulsion of Jews in 1492 was regrettable (Galanter 1992: 13).

It is ironic that what is arguably the least adaptive social institution in American history gained momentum partly because of the interaction of malaria as an environmental factor and what can be seen as an adaptive advantage in some West Africans -- the sickle-cell gene responsible for partial immunity to malaria and yellow fever. This immunity seemed to some to make West African slaves especially desirable in regions where European laborers suffered extreme sickness and mortality due to these diseases. Among those originally determined to live without slaves was a German-speaking community of Protestants founded in 1734 on the Georgia frontier. (They had been among those expelled by the Bishop of Salzburg in 1731.) The Salzburgers suffered terribly from malaria and yellow fever. Their leader, Pastor Johann Martin Boltzius "himself had a twenty year-bout with malaria, which weakened him as he got older" (Morgan 1993: 274). But although West Africans may have been better adapted than most Europeans to some aspects of life on the American frontier, the case can be made that the institution which brought them there was anti-adaptive. After decades of living by their own industry and despite the admonitions of Boltzius, some of the Salzburgers purchased slaves.

> As Boltzius had warned, slavery also helped wreck the colony, for, with slaves doing the work, the whole idea of the value of labor was lost by the others, and the settlement dwindled away. It became a ghost town. (ibid.: 275).

Morgan makes it clear that other factors contributed to the weakening of the Salzburger community as well, but it is clear that if we look at a social collective as an organism of sorts, then the adaptive value of the presence of West Africans as a kind of phenotype cannot be evaluated in any simple way. The community's fitness must be more than a laundry list of qualities; there must be an algorithm for evaluating those qualities *in combination.* In other words, we have to treat fitness in any context, including a social one, as conjunctive rather than disjunctive, but in a way bound up with theory, with *Weltanschauung.*

Probably there is no clear-cut way to deal with complicated synergies on a practical basis -- that was one moral of Meno's problem -- but let us ignore that for a moment. Just as challenging is the search for a yardstick of evaluation, but any yardstick we choose is liable to be theory-bound. Whole books apply this generalization to quantitative evaluation in general (e.g., Crosby 1996) in a way that boils down to making quantification dependent upon where we look and how we measure what we find. Carrier summarizes the two aspects: "Sometimes we are in need of a theory to direct our attention to those quantities whose investigation is to prove fruitful" and "...in measuring physical quantities one often has to rely on physical laws" (1994: xv, 9). But what applies to quantities goes for other scientific activities as well (e.g., building taxonomies, evaluating fitness). Ffirth reached the wrong conclusion because he designed his experiments and read his observations against what today's epidemiologists would take to be a false theoretical backdrop. Pastor Boltzius quarreled with slavery advocates among his neighbors and later in his own congregation because he measured the fitness of the Salzburger community by different criteria than his interlocutors.

Abstractly, such cases represent a problem for the taxonomy of multiple realizability which we have developed to this point. The ontology of our discussion has implied that multiply realizable contexts consist of *realizing* "basis" sets and *realized* phenomena. We will discuss the algorithmic "bridges" between basis sets and multiply realizable phenomena at greater length below. For the moment, suffice it to say that such bridges cannot exist in the absence of decision criteria. These yardsticks are not "given" by observation of raw nature nor by experiment in concocted settings; rather, they *precede* the gathering and interpretation of data. This is clear in the examples we have just considered. For instance we could say that Ffirth tested a true disjunctive basis set of two possibilities because he had not previously linked any other hypothetical causes to the transmission of malaria and yellow fever as a multiply realizable phenomenon. Similarly, sometimes the disconnect between anti- and pro-slavery spokesmen was a function of different ways of measuring slavery's "fitness" in the social collective. Small wonder that an advocate of self-sufficiency like Pastor Boltzius would fail to agree with those who did not take self-sufficiency into consideration.

It may be objected that such difficulties are irrelevant to most discussions of fitness simply because the interlocutors generally agree on their yardstick in advance. For instance, we may choose to measure real longevity and production of offspring, or we may turn to something like Mills and Beatty's (1979) propensity reading of these variables. But the variables themselves also admit of various interpretations. Do we want to measure fecundity in terms of raw numbers contributed to the next generation? We could, but biologists sometimes find that greatest litter size does *not* lead to the most surviving offspring, nor to the most progeny in second and higher-numbered generations (Morris 1986: 174, 178). We will discuss this issue further in a coming chapter; for now it suffices to remark that the context of multiple realizability always contains a number of theoretical assumptions whether that fact is explicit or not. As we will see later in the chapter, this has a definite effect on the reducibility of multiply realizable phenomena.

(5) Hypersufficiency

A further point of clarification is that the multiple realizability of fitness can present a conjunctive challenge even when the environmental context is constant. This problem, again a result of theory-laden perspective, can occur when a given organism apparently has *too many* adaptations to its environment. At first blush, that very notion seems to be a contradiction in terms. How can an animal possess too many adaptively advantageous traits in a given environment? Darwinian evolution is not like Aristotelian ethics, after all, in which "moral qualities are so constituted as to be destroyed by excess and by deficiency" (*Nicomachaen Ethics* II.2, 6; Rackham (tr) 1934: 77), and therefore we could dub fitness "twice-dimensional" above. In other words, we may be tempted to think that natural selection does not penalize "too much of a good thing" because that phrase has no meaning within the Darwinian system.

But here we must distinguish between actual traits and the criteria of inference used to determine how existing characteristics might correspond to ecological niches. Evolutionary biologists take seriously the possibility that an apparent excess of adaptation for a given lifestyle in a particular environment signals inferential error. Charig and Milner (1986), for instance, speculated that the "new" dinosaur *Baryonyx* may have been specially adapted to catch fish based on the structure of its snout, the nature of its teeth, and the morphology of its forelimbs. "It is suggested that *Baryonyx* crouched over the bank of a river and used its claws to hook fish out of the water rather like a grizzly bear" (Kitchener 1987: 114)[10]. But Kitchener is suspicious of this explanation of the observed phenotypes, in part because "[i]t seems to be that *Baryonyx* has too many adaptations for fish feeding. Why does it have both forelimbs for hooking fish, like a grizzly bear's, and teeth and jaws similar to the fish-eating gavial's, when one adaptation would suffice?" (ibid.) Kitchener prefers the theory that *Baryonyx* was a scavenger, because in that case "the need for both the forelimbs, to open up carcasses, and the narrow snout, to enter the body cavity, would be explained" (ibid.).

What we see in Kitchener's evaluation of *Baryonyx* is an odd pattern of reasoning but perhaps a valid one for all its apparent strangeness. If an abstract phenomenon can be concretely realized in more than one way, maybe it cannot be instantiated in all of them at once. In other words, although a number of linkages of the form general quality-to-concrete realization may appear possible, perhaps they all fail if too many are in fact possible. That is the surprising principle of Kitchener's argument. Apparently he would not have objected to the assertion that *Baryonyx* was a riverbank fisher if *either* the forelimbs were appropriate to that role (as they are, at least when viewed in isolation) *or* the nature of the snout and teeth were adapted to fishing (which, again, seems to be true). But the fact that each of these linkages is individually possible -- {claws implies fisher} *and* {snout implies fisher} -- rules out the possibility that the conjunctive linkage {Claws-plus-snout implies fisher}is correct. Or more precisely, the individual cases *tend* to rule out that possibility because organisms driven by the principle of natural selection *tend* toward *sufficiency* rather than *perfection*. That, at any rate, seems to be the basis of Kitchener's reasoning, whether we agree with it or not.

Let us call the principle of Kitchener's argument *hypersufficiency*. We will say that an organism or a collection of phenotypes is hypersufficient for a given ecological niche if the adaptations have been optimized beyond what we consider to be the requirements set by selective forces in the organism's environmental context. Note that hypersufficiency bears some resemblance to the self-bounding character of Greek virtues but does not destroy the twice-dimensional character of phenotypes, nor of Fitness itself. For although Kitchener finds it unlikely that *Baryonyx* developed to fit the niche of riverbank fisher, still he must admit that the dinosaur's traits *per se* are well adapted to that role.

We should note that Kitchener does not argue that either the facial structure or forelimb design of *Baryonyx* was a preadaptation, that is, a structure adapted to a purpose other than fishing from river banks which was then appropriated for that role in the way that a turtle's limbs may be adapted to swimming but used as well for digging nests in sand, or in the way that what we consider to be the speech apparatus of humans may have evolved for purposes other than producing speech:

> It has often been pointed out that what the linguist commonly refers to as the speech organs (or vocal organs) -- the lungs, the vocal cords, the teeth, the tongue, etc. -- all serve some biologically more basic function than that of producing vocal signals. This is indisputably the case: the lungs are used in breathing, the teeth in chewing food, and so on. None the less, all babies start babbling when they are a few months

> old...and babbling, which involves the production of a much wider range of sounds than may be found in the speech of those with whom the child comes into contact, cannot be satisfactorily explained in terms of the child's parrot-like imitation of the sounds that he hears around him. Furthermore, it has now been demonstrated experimentally that babies are capable, in the very first few weeks of life, of distinguishing speech sounds from other sounds and are predisposed, as it were, to pay attention to them. Man's nearest relatives among the higher primates, though they have much the same physiological apparatus, do not show the same predisposition to produce or distinguish the sounds characteristic of human speech. (Lyons 1981: 14 -15)

Why Kitchener does not entertain this possibility is not totally clear. Perhaps he believes that even appropriated phenotypes can be hypersufficient.

## (6) Summary of definitions and terminology

The discussion of multiple realizability and the terminology we have introduced to this point can be summed up as follows:

(1) The tentative definition of multiple realizability at (1.a) above -- "...a circumstance is multiply realizable if none of its (multiple) sufficient causes is necessary" -- is correct as far as it goes. However, three cases can be distinguished from one another:

(a) *True disjunctive* multiple realizability describes the case in which just one of several possible alternatives is the cause of a given multiply realizable phenomenon at a given time. A different phenomenon in the same genus (category) or the same phenomenon at a different time may be realized through another of the possible causes. We illustrated this case with the example involving possible evolutionary paths leading to the development of birds. Presumably there is a historical fact of the matter in this case, meaning that one of the paths is actually the route which the ancestors of birds traveled, so to speak; the other path, although conceptually possible, was in fact not taken.

(b) *Simultaneous disjunctive* multiple realizability is our name for the circumstance in which all of a number of possible causes can *simultaneously but independently* yield the same phenomenon. This was exemplified by what is often taken to be a classic case of multiple realizability or supervenience: distinct organisms -- even those belonging to different species -- can end up having the same fitness value (measured by criteria such as longevity and number of offspring produced) even though they have different phenotypes. The congruence relationship of various ratios to a given percentage value is arguably another example of simultaneous disjunctive multiple realizability.

(c) *Conjunctive* multiple realizability describes cases in which a phenomenon can be variously realized by different combinations of a basis set's elements. The possible combinations are thus sufficient but not necessary, whereas the elements within each realizing combination are not even sufficient. Again fitness serves to exemplify this kind of multiple realizability. Indeed, conjunctive realizability is probably a more accurate way than the other two kinds of explaining how different organisms can have the same fitness value, since individuals comprise numerous different phenotypes.

(3) Appealing to a Greek conception of the relationship between individual virtues and Virtue in the abstract helped us establish the convention of calling phenotypes "twice dimensional," since their quiddity as well as their duration can be optimized. The relationship also gave us Meno's problem -- the challenge of specifying precisely how the virtues are linked to Virtue or (in an imperfectly analogous way) how fitness values are associated with Fitness.

(4) A fairly traditional concept of the theory-ladenness of observation and experiment was given a "twist" to jibe with the foregoing discussion of multiple realizability. In a nutshell, our evaluation of multiply realizable phenomena is conditioned by our conceptual range. Such a desiccated, formal statement sounds innocuous and obvious, but in practice the dictum has far-reaching consequences, as Ffirth's experiments and conclusion demonstrate.

(5) Related to (3) and (4) is a pattern of reasoning exemplified by Kitchener's criticism of those who would hypothesize that the dinosaur *Baryonyx* fished from river banks. He bases his criticism on what we are calling the *hypersufficiency* of the organism's extant phenotypes for that role. Aristotle would have held it to be a contradiction in terms to say that someone had "too much courage," since courage is not variable along the quiddity axis. The hypersufficiency argument does not deny the twice-dimensionality of phenotypes, but it nonetheless subscribes to a theory of the mean reminiscent of Greek theories of ethics: if an organism is "too well adapted" to an ecological niche, then perhaps it is not adapted for that niche at all. By Kitchener's reckoning, *Baryonyx* had too many characteristics appropriate to the role of riverbank fisher. To make that kind of determination, we must of course pay

attention not just to the organism itself but also to the evolutionary impulse which controlled its phylogenetic development. This demonstrates a theory-ladenness which we observed throughout the discussion.

## 2. Applications

### (1) A practical benefit of multiple realizability: satisfying the "advantage clause"

Multiple realizability is frequently a matter of interest in biological investigations, and its presence holds benefits as well as challenges for the theorist. The utility of multiple realizability can be seen particularly in analyses concerned with the evolution of organisms and their various phenotypic characteristics, where it is necessary to speculate in order to explain how and why a given structure might have arisen. The multiplicity of conceivable causes can aid the evolutionist's agenda by suggesting possible selective advantages in structures which otherwise might seem utterly mysterious in their development or even contrary to a basic thesis of Darwinian evolution. We can call that tenet the "advantage clause": commitment to natural selection as evolution's motive force compels theorists to explain not just the utility of various existing behavioral and morphological characteristics; when such properties are complex, the developmental stages through which they developed must also be shown to have had a probable selective advantage, else it would be difficult to understand how the intermediate phenotypes were perpetuated long enough to lead to the successive and ultimate stages. (This matter differs from a related issue -- whether it is *possible* for a very complex structure such as an eye to have developed from simpler, non-eye-like precursor components. For commentary on the possibility, see Dewey 1910: 11; Dawkins 1995: 87 - 108.)

A challenge to entomologists, for instance, has been to explain the development of wings in insects. Apparently early wingless species evolved into winged ones, but how? Fossil evidence shows intermediate species which seem to have possessed "pro-wings," also called thoracic lobes -- structures which, although not wing-like in the modern sense, may have been a stage in the development of wings. Some entomologists have argued that these pro-wings arose as extensions of the thoracic terga (dorsal plates) and later became articulated, but that theory has been weakened by evidence suggesting articulated pro-wings may have converged with non-articulated tergic appendages. In other words, there is no clear evidence showing evolution starting from simple thoracic plates, moving on to thoracic lobes, and leading ultimately to the wings of modern insects. Such a progression might be more safely inferred if it could be shown that thoracic lobes, which were present in some species of Paleozoic insects but which could not have functioned as wings themselves, conform to the advantage clause -- that although they did not make flight possible, they did confer a different kind of fitness advantage on their possessors. This challenge has evoked a range of theories exemplifying the positive use of multiple realizability: "These broadly attached pro-wings ... may have functioned originally to cover the spiracular openings or gills in amphibious ancestors, to protect and conceal insects from predators, to facilitate passive aerial dispersal by small insects, or to aid in sexual displays" (Douglas 1981: 84; references to footnotes omitted). Yet another conceivable function of thoracic lobes is thermoregulation (ibid.).

### (2) The alleged hypersufficiency of single phenotypes

Attempts to reconstruct the history of evolution from the mists of multiple possible causes have led to many theories which follow the same pattern as Douglas's argument on behalf of the thermoregulatory benefits of pro-wings. At times an issue reminiscent of Kitchener's problem with calling *Baryonyx* a riverbank fisher arises, only in these cases, the alleged hypersufficiency inheres in a single quality of an organism rather than in a conjunction of phenotypes. Perhaps the paradigmatic case is "higher" intelligence. None would dispute that some instances of mental creativity confer survival advantage. The ability to construct the stone ax which some of my ancestors surely employed or the computer that I am currently using can be construed as huge leaps forward in the quest to earn our daily bread and stipendia, to fend off predators and typographical errors, to secure a warm, dry cave or a colder, damper student's apartment.

But what about poetry or music, products of human endeavor which may well have a practical communicative function but which seem also to possess elements superfluous to survival and reproduction? Of course the theorist can always play the sexual display "wild card" (recall that that explanation appeared in the list of possible utilities for thoracic lobes among pre-winged insects). This rhetorical move amounts to claiming that a phenotype for which no immediately practical survival advantage can be conceived must contribute directly or indirectly to sexual display, i.e., to attracting mating partners. Qualities such as a peacock's tail feathers or behaviors such as a ruffed grouse's "dance" are credible instances of phenotypes whose utility is indeed sexual display. But when credibility is challenged, such explanations are difficult to prove. This kind of challenge arises when sexual display arguments offend an intuition which whispers that the ability to perform a two-hour-long symphony or to write a five-hundred-page volume of poetry would be overkill if the sole utility of these capacities were to arouse the opposite sex. Intuitively, we find tabloids credible when they document the mating successes and excesses of moderately talented rock-and-rollers even as we read rumors that Leonard Bernstein ended up in bed with the likes of other composers such as Ned Rorem and Aaron Copland (Peyser 1987: 159). Thus we can speculate in a rough and ready way that even if basic musical competence is adaptive as a means of sexual display, musical genius may not be similarly advantageous, and the same judgment applies to the cognitive resources underlying the respective levels of musical endowment. At some point mental capacities and the talents which they make possible seem superfluous to the business of surviving and reproducing.

Human intelligence is therefore problematic for reasons reminiscent of the hypersufficiency argument exemplified by Kitchener's opinion on the ecological role played by *Baryonyx*. But whereas Kitchener could imagine a niche in which the dinosaur's attributes would be properly sufficient, certain talents and their mental foundations have been seen as hypersufficient for *any* role whose goals are mere survival and reproduction. Human intelligence even opened a philosophical rift between Darwin and Wallace (Gould 1977: 50; 1980: 47 - 58; Restak 1979: 89), a difference of outlook which potentially affects how we evaluate the Darwinian theory of evolution overall. The value of scientific theories is appraised partly based on their generality: the more extensive the explanatory bailiwick of a given theory, the more powerful that theory is held to be. The genius of Darwin's identification of natural selection as the motive force of evolution lies partly in the theory's applicability to the entire *scala naturae*. In our day, most sympathetic readings of Darwinian evolution --

including essentially theistic ones (e.g., Teilhard de Chardin's, 1956, 1957; cf. Barrow and Tipler 1988: 195 - 205[11]) -- hold that the development and present status of all living things, not excluding human beings, can be explained in one fell swoop. On this account, there is no need to attribute different agencies to various phenotypes and behaviors; all characteristics of animals are explained by the pressure to adapt to environmental conditions (though religious and non-religious evolutionists may handle the transition from non-living to living matter differently, which can be seen as amounting to a disagreement about the boundary of the *scala*). One obvious exception to this pattern is appeal to random walk as the primary engine of change. Another exception is that some theorists who hold themselves to be consistent Darwinians believe phenotypes can be "forced" by other elements of a given organism's *Bauplan*, and thus not all phenotypes are the direct results of adaptation. This issue will be discussed at some length below, so for now it is only necessary to point out that those who buy into the *Bauplan* argument are not really asserting the existence of an agency other than natural selection so long as the *Bauplan* itself is held to be adaptive in its origin. Thoroughgoing adaptationists as well as those we might call "bauplanists" can therefore be seen as consistently Darwinian.

Wallace co-founded and thus certainly endorsed the theory of evolution by natural selection, but he threatened to reduce the theory's range by excluding human beings (Gould 1977: 50; 1980: 47 - 58). His reasoning was based on what he perceived to be (to use the terminology which engages us at present) the irreducibility of human intelligence and its resulting capacities to a set of adaptive forerunner capabilities. If we were to trace the origin of human intelligence as a phenotype backward in time, we would find that its near predecessors also could not have fulfilled the advantage clause. For Wallace, the superfluity of musical ability compared with the apparent survival and reproductive challenges of the environment -- what we are calling the hypersufficiency of musical ability -- sufficed to show that humans stand at least somewhat outside the realm in which evolution proceeds by natural selection. Darwin adopted what we would consider a more modern outlook, optimistically maintaining that however far removed from the struggle for survival such mental endowments and their predecessor phenotypes may appear to be, all can ultimately be explained within the framework of evolution by natural selection.

At present there seem indeed to be credible reductive explanations for musical ability and for intelligence in general which do not rely on speculations about sexual display and which do not push animals possessing "higher" intelligence outside the Darwinian schema. Neurophysiologist William H. Calvin suggests that musical and even linguistic capacities may stem historically from an ability which has clear adaptive value.

> "As improbable as the idea initially seems, the brain's planning of ballistic movements may have once promoted language, music and intelligence. Ballistic movements are extremely rapid actions of the limbs that, once initiated, cannot be modified. Striking a nail with a hammer is an example" (Calvin 1994: 104).

It appears that the cerebral prerequisite of such movements is the ability to plan in advance, very precisely, an ordered sequence of muscular movements. The need for extreme accuracy in planning is clear when we reflect that once a ballistic movement has begun there is no time for correction, nor even for the reception and processing of sensory feedback on which corrective action would have to based. The ability to produce such precise and complex plans prior to ballistic movement is easily linked to intelligence in general, and specifically to the kinds of mental capacity necessary to produce music and perhaps a range of other products whose immediate survival and reproductive advantages are not immediately clear.

Once again it is apparent that to the extent mental abilities are adaptive for the organism, they are so only in combination with other phenotypes. Put another way, although it may be convenient to speak of the fitness of *phenotypes* as being multiply realizable in a simultaneously disjunctive way, the fitness of *organisms* must always be reckoned conjunctively. In general we do not think of fitness as having a true disjunctive character, either as an attribute of phenotypes or of organisms. That is, we tend *not* to think that a given fitness value can result from *either* one realizing basis, *or* another realizing basis, *but not both.*

### (3) The diachronic character of perceived multiple realizability

Apparently there are no *a priori* grounds for ruling out two possibilities. The first possibility is that there could exist a rare phenotype which, in combination with mundane attributes, would lend such a high fitness value that no other combinations excluding the Super Phenotype could hope to compete by yielding Super Fitness. The second possibility is that a particular combination of normal phenotypes would yield the same Super Fitness. For instance, we might adapt Wallace's position on human intelligence to make the claim that a specific phenotype (if "higher intelligence" could be considered as such) affords a level of fitness which is *sui generis* among all fitness values -- there is only one way to reach it, even though we might have taken other combinations as realistic possibilities and ruled them out only through observation and experiment. In that case, a super-high fitness value might possess the character of what we have called true disjunctive multiple realizability (only one phenotype or one combination of phenotypes would afford the possessing organism Super Fitness).

In this scenario there would seem to be several possibilities prior to investigation, but after all have been weighed and tested, only one "path" to that value would actually be possible. Such a case differs from our earlier example of true disjunctive multiple realizability, the case involving alternative evolutionary paths leading to the emergence of birds. In that instance, we considered that there were two possible facts of the matter, cladistically speaking,[12] but from the outset we knew (or at least strongly suspected) that the two options were mutually exclusive. We should take notice, however, that in all such instances -- in our example of alternative phylogenies of birds as well as the possibilities of a Super Phenotype or a Super Combinations -- the multiple realizability has a diachronic character. In other words, *after* we know the fact of the matter, there is no longer any multiple realizability. Once we know the real phylogeny of birds, the other path is discarded. If a Super Phenotype exists, then either its corresponding fitness value is truly one of a kind (hence there is no multiple realizability) or else it realizes a fitness value which can also be effected through other means (i.e., the multiple realizability becomes simultaneously disjunctive).

From these speculations we can conclude that true disjunctive multiple realizability depends upon a diachronic aspect. We could even call it "historical" multiple realizability, in that the multiple "paths" leading to the single result in question are only tentatively possible, before we have the opportunity to gather data and to formulate and test hypotheses. The cladistic niche of birds is multiply realizable until we know the historical fact of the matter; a Super Phenotype is either not multiply realizable or else it can be part of a simultaneous disjunctive set, not of a truly disjunctive one.

### (4) Quiddity, not just quantity

To this point we have dealt with multiple realizability by considering what could be called "order relationships" between sets of basis circumstances and the multiply realizable phenomena which those circumstances, alone or in combination, can effect. (In algebra, the "order" of a group is the number of elements it comprises, q.v. Herstein 1975: 28.)[13] Our primary interest has been in examining variations of the many-to-one relationship which characterizes multiple realizability. But of course the elements of the realizing set have their own quiddities, as do multiply realizable phenomena. To the extent that the abstract phenomenon of multiple realizability has to do with explanation or definition of a given situation *in terms of* or *with respect to* sets of realizing possibilities, paying attention to order alone is insufficient. We have to look at the nature of the elements involved in both sides of the many-to-one relationship, as well. The question is whether there are any generalizations which we can make. After all, we think of multiple realizability as expressing the many-to-one relationship existing between phenomena in many different fields. We cannot hope to look at all cases of multiple realizability, of which there presumably are uncountably many. Rather we must hope to find abstract categories dealing with *quiddity* just as we have thus far dealt mainly with *quantity* (order). That is the business of the next section.

## 3. A first glimpse of recursion: layered multiple realizability

With little effort we could find many more examples of multiple realizability in evolutionary biology, but for the moment the foregoing illustrations should suffice to show the various quantitative aspects of the concept which will be important later in our discussion of recursive fitness. What is lacking in the account thus far developed is a closer examination of the two "sets" involved in instances of multiple realizability -- the "basis" realizing properties and the phenomenon which the basis realizes. Most of the examples offered to this point have treated the two categories, the realizing and the realized, as being essentially heterogeneous. The basic pattern has gone something like this: we have a number of concrete phenomena which, singly or in combination, suffice for (or *cause*) the existence of another phenomenon. The thing which is thus realized does not belong to the set of base phenomena, however, nor is it equivalent with a combination of all of them in the case of conjunctive multiple realizability.

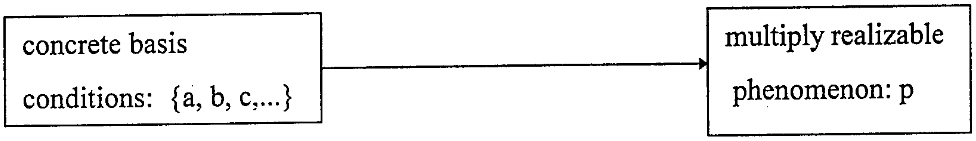
To emphasize this point, let us quickly run through a few of the examples already employed. We began by considering Darwin's letter to Gray, in which the Englishman pointed out the irreducibility of percentage values to ratios. It is clear, for instance, that "2/100" and "1/50" suffice to realize "two percent." Now in this case there appear to be an uncountable number of sufficient "bridges" linking the realized phenomenon with realizing conditions in a basis set, but none of the bridges is necessary. Let us assume that a sufficient bridge is a kind of *function* in Frege's 1891 sense (*Funktion*), with ratios such as 2/100 being *arguments* of the function and two percent being its *value*. (The function here is simple: it normalizes the denominator of a ratio to 100 while simultaneously adjusting the numerator by the same proportion, then identifies a percentage by the value of the numerator.) Apparently we have discerned three separate kinds of entity involved with bridging -- argument, function, and value -- none of which can be made literally equivalent to the others despite the conveniences of everyday speech (the inherent error which Frege endeavored to correct). Nonetheless, a nagging sense may remain that even if no biconditional (necessary and sufficient) relationship unites two percent and 2/100, the two are somehow very similar.

What "very similar" amounts to in this case is unclear, but certainly the generic likeness seems intuitively greater than that between, say, the sensation I feel when a cat brushes up against my leg and the cat itself. The same holds for many of the other examples we have used: there is a set of base phenomena, one or some or all of which (depending on whether the instance of multiple realizability is truly disjunctive, simultaneously disjunctive, or conjunctive) act to realize the multiply realizable phenomenon. We have seen, for instance, the following relationships among others:

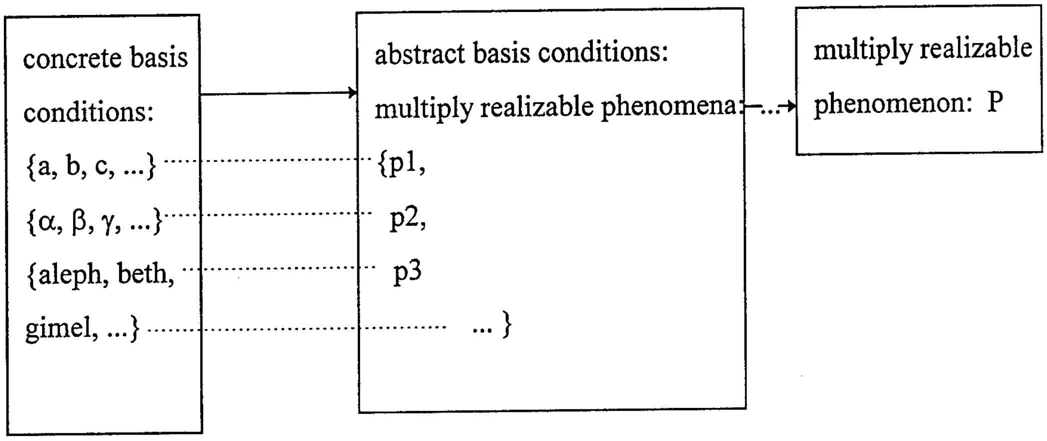
| | **Basis Set** | **Phenomenon** |
|---|---|---|
| (1) | {cat, wind, dog, illusion, ...} | sensation on my leg |
| (2) | {organism$_1$, organism$_2$, ...} | single fitness value |
| (3) | {evolutionary path$_1$, evolutionary path$_2$, ...} | birds |
| (4) | {long neck, long tail, ...} | fitness values of giraffes |
| (5) | {bodily contact, bad air, ...} | transmission of malaria |
| (6) | {justice, temperance, piety, ...} | Virtue |
| (7) | {phenotype$_1$, phenotype$_2$, ...} | Fitness |

The argument could be made that while examples (1) through (5) are clearly cases of heterogeneous multiple realizability, examples (6) and (7) possess that nagging kind of similarity which also tickled the intuition in the multiple realizability of percentage values. (Of course there is a counter-argument to this position, as we can see when we compare, for instance, examples (4) and (7). But that is beside the point for the moment. Our present interest is in trying to discover whether the vague and intuitive distinction between homogeneous and heterogeneous multiple realizabilities can be made more rigorous.)

If a homogeneous type of multiple realizability can be distinguished from a heterogeneous kind, perhaps the distinction will rest on the difference between *explaining* and *defining*. In all of the examples noted above, it is clear that we can *explain* the existence of the realized phenomenon by appealing to members of the realizing basis. We can draw a causal connection between a sensation (what I feel on my leg) and various environmental conditions (cat, wind, etc.), for instance. (It is not too great a stretch to include an "internal" condition such as tendency to hallucinate as a part of the "environment" in this instance.) However, we cannot wholly *define* the sensation in terms of those same basis conditions. Of course we can say something like "That sensation *is defined as* what one would feel if a cat were to brush up against one's leg, or if a strong wind blew across one's skin, etc." Although there may be a place for such *causal* definitions, as a category they do not exhaust what we consider to constitute definitions in general. Apparently some definitions rely on analogy -- they function by saying what a *definiendum* is *like*. This is not the place to attempt an exhaustive theory of definitions, so let us instead consider two patterns:



direct multiple realizability



layered multiple realizability

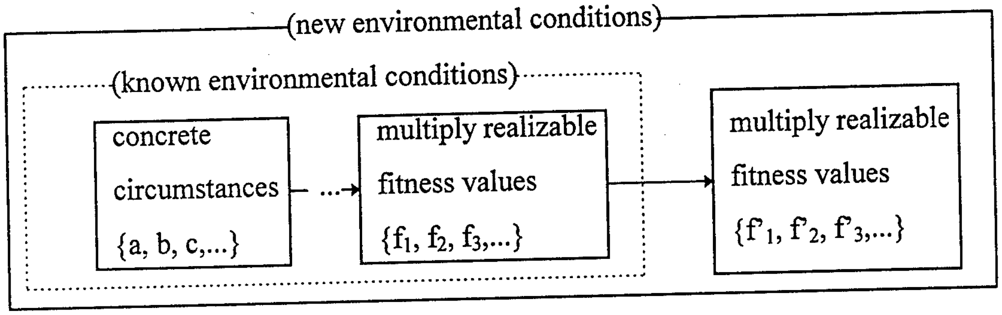
These diagrams attempt to show that the intuitive similarity between the realizing basis and the realized phenomenon in some of our examples has to do with the fact that the realizing conditions of some multiply realizable phenomena can themselves be multiply realizable. Take the relationship between ratios and percentages or of virtues to Virtue, for instance. In such linkages there is a degree of similarity between the elements of the realizing basis set and the multiply realized phenomenon simply because both "sides" of the relationship can be seen as multiply realizable; one side is simply further removed from concrete circumstance than the other. A proportion is multiply realizable just as a percentage is (obviously, since the very terms *pour cent* or *pro centum* are clearly particular kinds of proportion). Similarly, it appears that there are certain concrete conditions upon which an individual virtue such as piety can be realized; piety, in turn, becomes one of the realizing conditions of Virtue in the abstract, at least under some conceptions. Perhaps the same can be said of specific fitness values: arguably a fitness value at a certain level -- "sub-fitness," we might term such a concept -- is realized by concrete circumstances (the morphology and behavior of an organism in the context of its environment); in turn, a concept of Fitness in the abstract or what we might call "Super Fitness" exists at a higher level, metaphorically speaking, and encompasses the concepts of sub-fitnesses by being realized upon them. (Conceivably the layering could be much deeper, but for the moment it suffices to consider the simplest case for explanatory purposes.) Perhaps the intuitive similarity of multiply realizable phenomena at different levels in the same continuum can be explained by noting that the elements at all higher levels are, ultimately, *functions* of the concrete basis conditions. The higher one goes, the further removed one is from that concrete foundation, but the pattern of the architectonics can always be reflected as a relationship between the current level $L_i$ and the preceding levels all the way down to the basis B. Looking from the ground up, so to speak, we see a progression of functions:

$$L_1 = f_1(B),\ L_2 = f_2(L_1),\ \ldots,\ L_i = f_i(L_{i-1}).$$

(It should be noted that these expressions treat the functions involved as *potentially* but not necessarily different at each level; the case that $f_i = f_{i-1}$ is not excluded for any or all values of i.)

The question of whether phenotypes can be identified along non-arbitrary lines cannot be answered here, but let us suppose that we have an exhaustive list of the characteristics of various organisms. In addition, assume that all environmental factors are known and that we have some means of accommodating "common" random interactions of organism and environment (e.g., fatal lightning strikes). Finally, suppose we have track records of longevity and reproduction for identical organisms in the same environment. Under these epistemological conditions we might calculate a tentative fitness value for each organism. Now we can conceive the possibility that our knowledge of organisms and environment might be incomplete. Maybe some new environmental condition -- caused by the eruption of a volcano, for instance -- suddenly arises. In this circumstance we would not wish to begin reckoning the fitness of organisms from scratch; rather, we would attempt to "piggyback" on the knowledge we already possessed. (Whether our extrapolations would be successful is another matter.)

Diagramatically, the case could be represented as follows:


(new environmental conditions)
(known environmental conditions)
concrete circumstances $\{a, b, c,\ldots\}$
multiply realizable fitness values $\{f_1, f_2, f_3,\ldots\}$
multiply realizable fitness values $\{f'_1, f'_2, f'_3,\ldots\}$

In such a scenario, the fitness values computed for the new environment will be a function of those which applied in the old one. Depending on how dramatic the environmental changes actually are, the new fitness values will differ little, if at all, from the old ones.

## 4. Chapter summary

The taxonomy of multiple realizability which we have constructed may not be exhaustive, but it contains important distinctions which will be exploited in upcoming chapters. We have seen that some phenomena are multiply realizable in significantly different ways (including *truly disjunctive*, *simultaneously disjunctive*, and *conjunctive* manners). An issue not resolved is what we have called "Meno's problem," coining the term from the character in Plato's exposition of an unsolved question: How do we express the relationship of a whole to its parts in synergistic conjunctions? In Plato's discussion the question was couched in terms of the relationship between virtues and Virtue; for us, the issue is how "layers" of fitness relate to other layers and to concrete circumstances.

It was further noted that some basis sets are what we have called *hypersufficient* to realize a given phenomenon when a certain motive force is stipulated as the function which bridges between the basis set of realizing phenomena on the one hand and the realized phenomenon on the other. The discussion led us to a tripartite view of multiple realizability, involving a phenomenon, a foundational set (basis) of possible ways to realize that phenomenon, and a function or algorithm which constitutes the means of bridging between the realizing and the realized. By looking more closely at the way in which these three elements interact, we discovered that multiple realizability can be "layered" in such a way that for a given multiply realizable phenomenon, the most proximate basis set of realizing possibilities may itself be multiply realizable. This regression of multiply realizable entities partly accounts for an intuitive sense that some multiply realizable contexts are more nearly homogeneous than others. The next chapter attempts to refine the sense in which layering can take place in multiply realizable contexts by introducing the notion of recursion.

# Chapter Three: Recursion Defined

It was suggested in the introduction that the concept of recursion provides a new way of evaluating problems, somewhat in the same way as the calculus allowed approaches which had not been possible before Newton and Leibniz. It was not that one could not have spoken of a slope or of the area under a curve prior to that time; on the contrary, such conceptions seem to arise naturally, perhaps even reflexively, from observing a curve. But by providing a *method* of calculation, the calculus makes possible more abstract conceptions which might not exist otherwise. This is an intriguing phenomenon (assuming it in fact exists): quiddity arising from method. Consider a notion such as a third-derivative applied in a certain context -- say, the third derivative of a curve plotting distance traveled against time. The first two derivatives we recognize easily as velocity and acceleration, and these are "things" which correspond to the vocabulary of common discourse. But how do we describe the third derivative? Arguably "the rate of change of acceleration" is associated with a statement such as "They started going faster and faster," assuming that the movement described is not just one involving acceleration but a climbing rate of acceleration. Somewhere, though, commonly used words and phrases will utterly fail to describe something we can conceive based upon a method of abstraction. To be safe, let's say the tenth derivative of a curve has no analog in common parlance. Such a notion piggybacks on prior concepts. You don't know what the tenth derivative is? Well, I'll tell you: It's the derivative of the ninth derivative. Don't know what the ninth derivative is? No problem. It's just the derivative of the eighth derivative..., and so on, until you understand that the tenth derivative is the value of a function which we get by taking what was the function's value and making that prior value the argument of the same function, and doing that a certain number of times.

To keep things simple, let's go back to the third derivative. We're calling it the rate of change of the rate of change of the rate of change of the curve at a given point. In a natural language, identical phrases thus concatenated can be more confusing than enlightening. It helps to use some sort of symbol to demarcate the units of analysis -- perhaps parentheses yielding something like "change(change(change)))" in this case. Obviously there is a kind of self-reference going on here, but it will not be argued in this dissertation that the calculus requires self-reference. What engages us in this chapter is the idea that the kind of self-reference called recursion is analogous to the calculus in that both infuse our treatment of old problems with new quiddity borne of new method. The next several pages provide a bare-bones "toolbox" of concepts by discussing recursion as object and method. The two following chapters focus on the application of these insights to other key terms in our treatment of fitness as a recursive phenomenon.

## 1. A first definition of recursion

The father of the computer programming language Pascal, Niklaus Wirth, offers this simple definition in a textbook chapter entitled "Recursive Algorithms" (Wirth 1986: 135 - 170; cf. Frenzel 1987: 234):

(D1) "An object is said to be *recursive*, if it partially consists or is defined in terms of itself" (ibid.: 135).

Wirth quickly turns to concrete examples without formally elaborating on the difference between what might be called (on the basis of the definition above) "partially-consisting recursion" and "self-referential recursion." That will be our task below. In the course of the discussion we will need to clarify the applicability of the qualifier "recursive." The motive of this inquiry can be encapsulated in a single question: What are we to make of a chapter such as Wirth's, which styles itself a treatment of recursive *algorithms* but whose central definition is of recursive *objects*? That surprising conflation of method and object can be observed not just in the tension between Wirth's chapter title and the locus of the chapter's core definition of recursion, but indeed throughout his discussion. Understanding the two basic senses of recursion -- as object and as method -- will be one of our most important goals.

## 2. "Partially-consisting recursion"

We can divide definition (D1) above into two "halves." The first reads:

(D1.a) "An object is said to be recursive, if it partially consists of ... itself" (Wirth, ibid.).

The salient phrase is "partially consists of ... itself," and there seem to be two ways to understand what Wirth means. The first (trivial) understanding is to translate the phrase into something like "consists of parts of itself." This reading must be rejected on the grounds that it does not distinguish a recursive object from any other sort, since all composite objects consist of parts of themselves. We must therefore seek another way of understanding definition (D1.a).

### (1) Recursion as a diachronic process

If we ask what it means for something to consist of itself, it seems natural to apply a distinction from basic set theory. We say that {a} and {b} are *proper* subsets of the set {a, b}, whereas {a, b} is an *improper* subset of itself (Herstein 1975: 2). To put the distinction a bit differently, an object can be said to consist either of an improper subset of itself (i.e., an object *is* itself), or else the object can be described in terms of two or more proper subsets of itself (i.e., an object comprises *all* of its parts). These two possibilities seem to be exhaustive. Wirth's definition of what we are calling "partially consisting" recursion (D1.a), however, must be based on a third alternative, one which potentially violates the basic definition of proper and improper subsets. By saying that a recursive object consists *partially* of itself, perhaps he means that an object can be a proper subset of itself. But that is a seeming impossibility! (If we accept the common understanding of the law of excluded middle (Angeles 1981: 153), then a thing either is or is not itself.) At first glance, in short, there would seem to be no chance that an object can be recursive in the "partially-consisting" way which the first half of Wirth's fundamental definition implies.

However, that conclusion assumes the recursive object is (and therefore can be) *synchronically* present in its entirety. If an object is defined *diachronically*, by contrast -- if it "unfolds" across time, so to speak -- then it might be possible to interpret definition (D1.a) as claiming something like this:

(D2) An object O: $\{o_1, o_2, \ldots\}$ can be recursive if at any time $t_i$, the extant subset of O is different than that at another time $t_n$. Both subsets, however, are necessarily parts of O in their respective time frames.

(It is probably obvious that this definition says "can be recursive" rather than "is recursive" because, as it will be recalled, Wirth's definition has another "half" besides the one referring to what we are calling "partially-consisting recursion.") At time $t_1$, for instance, object O might consist of the subset $\{o_1, o_3\}$, while at $t_2$ O might be constituted by $\{o_3, o_4, o_5\}$. This still does not explain how the object can "consist partially of itself," however. To do that we need some further rhetorical tactic. If we apply Frege's distinction between argument, function and value, we can treat an object (set), *understood as a value*, to be yielded by a function defined on that same set *understood as an argument* (Frege 1891). Now suppose that the function is defined such that its *value* at $t_n$ always becomes its *argument* at $t_{n+1}$. (Arguably a function cannot determine its own argument. If that is true, then the state of affairs described actually requires at least *two* functions, one which calculates S and one which assigns the argument for the next iteration. However, that detail need not concern us here.) Then to say S "partially consists of itself" could mean that a later value of S, $S_{n+1}$, always incorporates an earlier value of S, $S_n$, as argument.

An *a fortiori* approach to this hermeneutic problem would be to argue that the existence of things in general is somehow a function of time. This position could be understood not just in the sense that many or arguably even all objects exist across time (an ontological perspective) but also with respect to how we understand and talk about such time-bound entities (an epistemological or psychological outlook). To mention a figure such as Heidegger in this respect is to run the risk of being superficial unless six or seven hundred pages space are subsequently devoted to a fuller exegesis, but let us accept that risk while promising ourselves to look more deeply into historicity in the reckoning of fitness later on.

First it needs to be said that for Heidegger the notion of temporality has a significance which reaches far beyond our present purpose of analyzing recursion as applied to the concept of fitness. For instance, Heidegger seems to hold that the perception of temporality results from the proper acceptance of one's own mortality (the state of mind he called *Eigentlichkeit*). This is a fascinating inversion of the notion that time's passage is a *memento mori*, but it is relevant to us only abstractly, in so far as it implies that temporality is not properly perceived until one has considered the problem of identity. Fortunately Heidegger provides us insights which are more directly related to the problem of how being and time are related. That, after all, is essentially the challenge of the definition of recursion: What does it mean for a recursive object to be partially itself rather than, simply, itself? The answer provided by (D2) above amounts to "temporalizing" the copula in (D1.a): object O is sometimes known as one subset of the identities it assumes in the fullness of time, while at other times it is associated with other subsets of its overall being, at least in the way we perceive and talk about it. The unity of the object over time emerges from the fact that it always takes an earlier instantiation of itself as argument and is always reckoned by the same function. This seems to be roughly what Heidegger had in mind when he asserted (1927 II.4.68: 349; my single, Heidegger's double quotation marks; cf. Maquarrie and Robinson (tr) 400 - 401):

> But because speech is always discussion of the existing [that which exists], although not primarily and predominantly in the sense of theoretical assertion, the analysis of the temporal constitution of speech and the explication of the temporal characteristics of language structures can be started only if the problem of the fundamental connection of being and truth is 'unpacked' from the perspective of the problematic of temporality. Then one can delineate the ontological meaning of the "is," which a superficial [äußerliche] theory of proposition and judgment has deformed to a "copula". Only with respect to the temporality of speech, that is, of *Dasein* in general, can the "origin" of "meaning" and the possibility of a formation of concept [Begriffsbildung] be made ontologically intelligible.

(I take it that Heidegger's reason for thinking of the "is" as other than a copula matches Frege's rationale for insisting that a function and its value are not truly equivalent.) This perspective is complemented by another of Heidegger's views -- that Greek philosophy, including that of Plato and Aristotle, uncritically equated being with "permanence of presence" and thereby diverted attention from an essential and unsolved question in philosophy.

Needless to say, Heidegger does not discuss recursion *per se*. But his dual notion that objects have their being only in time and that the existence of an object can be analyzed and described only in time raises an issue which a definition such as Wirth's (D1) might allow us to overlook. The matter has to do with the distinction between recursive object versus recursive method. Apparently if we speak of a recursive object then the method of fragmentation employed in (D2) is based on the temporally composite character of the object itself. In other words, we can (paradoxically) treat an object as being a proper subset of itself if the object changes over time. We take new subsets into consideration as they arise and as our purposes of analysis dictate, but strictly speaking, this is not a matter of convenience. Rather, the object *is* self-referentially and therefore temporally composite. Heidegger seems to buy into that sort of ontology -- one in which we use names implying the existential continuity of objects only as a conceptual convenience, since being is essentially time-bound. A single "thing" is so only until it changes into something else.

But we can also understand recursion as a *method* in which an object or circumstance is held to be essentially static across time (whether it is or not) because an algorithm associated with it functions as a kind of referential glue. The thing under consideration can then be analyzed in parts, first one subset of its qualities and then another, without losing its identity because it also has an algorithmic character and the algorithm remains stable. From the passage quoted above it is clear that Heidegger views analysis and its tool, language, as inevitably proceeding in a serial fashion, although his outlook does not necessitate the sequential reexamination of the same object, using the same algorithm of investigation, which we associate with recursive method as defined by Wirth. Let us examine the tension between recursive object and method more closely.

## (2) Recursion as object and method

It is apparently easy to ignore the fundamental distinction between recursive objects and recursive methods. Wirth, for example, entitles his chapter "Recursive Algorithms." Then, as we have already seen, he provides an explicit definition of a recursive object. Shortly thereafter he offers a section entitled "When not to use recursion" (1986: 138 - 140) -- aimed, of course, at students of programming *techniques* -- which uses "recursive" as a descriptor of method rather than object. For Wirth's purposes this unacknowledged dual usage is not amiss. It may even aid his agenda of communicating the practical aspects of recursive programming in the shortest possible space. Or one might say that the mixing of object and method exemplifies the modern outlook (characterized by Lachterman 1989) which conflates object and method in analyzing mathematical "things" in general.

Wirth suggests that the "power of recursion evidently lies in the possibility of defining an infinite set of objects by a finite statement" (Wirth 1986: 136). The act of defining needs to be understood as not just analytical (exploratory) in this context; to define recursively in Wirth's sense is also to *create*. The temporally unfolding objects in his infinite set are brought into being by the algorithmic manipulation of previous subsets of the same set.

Definition (D2) treats the identity of a recursive object as a cumulative property instead of something which can be exhaustively mapped by observing the thing at just one moment in time. Metaphorically, one might conceive such an object to be represented by a movie as opposed to a snapshot or a single still-frame. All of the tens of thousands of frames a VCR tape can be considered "atoms" of the motion picture as a whole, but no single frame is the movie. To put the same idea in slightly different language: each frame is arguably necessary to a specific film but none is sufficient for that particular film, or looking at the matter from another perspective, some collection of frames is sufficient for a film in general but no collection of specific frames is necessary for a film in general. (The qualifier "arguably" is here meant to take into account a possible objection. It might be claimed that no single frame of, say, the film "Casablanca" is necessary, since if we were to remove a second or two worth of celluloid the movie would still be recognizable. This observation need not concern us, since recognizability in a coarse sense is not the central issue here. Our concern is rather identity. We want to know what it is that causes something to be itself. By careful comparison we could distinguish the edited copy from an undamaged version of "Casablanca"; therefore, the two films cannot be considered identical.)

The film analogy takes us only so far, since it would be difficult to argue that a movie matches our heuristic sense of what it is to be recursive. Our intuition in this regard surely includes the quality of self-repetition, but the average film does not repeat itself in any significant way. Quite the contrary, we generally go to the theater expecting to see a plot played out on the screen, and what we call a plot usually comprises a distinct beginning, middle, and end, something in which the sequence of the parts is as central to the meaning of what we see as the content of the individual "moments" of the film. It should be noted, however, that Wirth's definition does not preclude a critical role for inherent sequentiality. In fact, a computer programmer would find little use for a recursive algorithm which "calls" itself independently of other conditions. Where the average motion picture fails as a metaphor for recursive object in Wirth's sense is in the content of the individual parts. In general such frames or even much longer segments of film do not mirror the progress of the movie as a whole.[14]

The opposite is true of recursive objects. They are atomized in the sense that they are somehow incomplete at any given point in time, but on the other hand, every instantiation of a recursive object mirrors the object as a whole in some way. That, at any rate, is a reasonable heuristic interpretation of Wirth's exposition of recursion. For instance, he offers a caricature of recursive objects in the form of a question: "Who has never seen an advertising picture which contains itself?" (Wirth 1986: 135). Other "popular" presentations of recursion rely on the same sort of image. Hofstadter suggests that recursion might be represented by "Russian dolls inside Russian dolls" (1979: 127). In fact, the sense in which recursive objects partially consist of themselves can be rephrased in another of Hofstadter's pithy summaries: "...a recursive definition never defines something in terms of itself, but always in terms of simpler versions of itself" (ibid.). This is a rephrasing in so far as it restates our conclusion that proper rather than improper subsets are the building blocks of recursive objects -- simpler objects precede more complex ones.

Or is it a rephrasing? Is a proper subset a simpler version of the set as a whole? Not necessarily, since a subset can be very different from the set which it helps constitute, just as a single frame of a two-hour motion picture differs from the film considered in its entirety. That makes Hofstadter's phrase "simpler versions of itself" unsettling. Above we encountered the problem of how something could "partially consist of itself," that is, be a proper subset of itself. We tried to solve the paradox by defining the recursive object diachronically (definition (D2) above), but that tactic seems to fail in the face of Hofstadter's phrase, since proper subsets in general are not simpler versions of the sets in which they are contained. That is true regardless of temporal considerations.

We are left searching for a way to make clear how recursive objects differ from simple composite objects such as movies, which can be said to conform to Wirth's definition (D1.a) because they partially consist of themselves but which fail to fulfill Hofstadter's heuristic criterion of self-representation in their parts. In this case, it would be more proper to say that recursive objects consist of parts of themselves. But that is a loose way of expressing an intuition about recursion; it is not a rigorous translation of Wirth's definition (D1.a). Perhaps the solution to our problem depends on the concept of cumulativity, which we can incorporate into a refined definition based on (D2):

(D3) An object $O_i$: $\{o_1, o_2, \ldots\}$ can be recursive if at any time $t_i$, $O_i$ consists of all sets $O_n$ at $t_n$ where $n < i$.

In other words, a recursive object can consist of all the past iterations of itself plus whatever is new in the current iteration.

The definitions at (D2) and (D3) employ proper rather than improper subsets to explain how an object can "partially consist" of itself, in accordance with the first half of Wirth's definition of recursion (D1.a). That is important because it means we can take the qualifier "partially" quite literally, consistent with its everyday usage; no logical or mathematical tricks need be played with the notion of set *per se*. If there is any legerdemain involved here, it is the introduction of time into a question which seemed at first glance to have to do only with existence. The challenge, then, was to make credible the introduction of a temporal element in definition (D2). Definition (D3) shows how cumulativity (which entails a temporal element) functions in an explanation of recursion.

### (3) Examples linking sequentiality and closure in recursive structures

If we seek an example to make this breed of objects concrete, we require something whose identity has a cumulatively self-repetitive character. Small wonder that the average movie fails to fill the bill. Here we can turn to Wirth's further exposition of recursion. He notes that "recursion is a particularly powerful technique in mathematical definitions" (ibid.) and goes on to offer three examples (ibid.: 135 - 6).

| 1. Natural numbers: | 2. Tree structures | 3. The factorial function n! (for non-negative integers): |
|---|---|---|
| (a) 0 is a natural number | (a) O is a tree (called the empty tree). | (a) $0! = 1$ |
| (b) the successor of a natural number is a natural number. | (b) If t1 and t2 are trees, then the structures consisting of a node with two descendant trees is also a tree. | (b) $n > 0$: $n! = n * (n-1)!$ |

In the first of these three examples, it seems clear that we would define a given number's successor in terms of cumulativity. Here, the cumulative character happens to be literal. In fact it could be considered a kind of additivity, since the successor of a natural number is that number plus one. Since the discussion to this point has also stressed the temporal nature of recursive objects, we should consider in what sense the first example is dependent upon time. Of course it can be argued that all natural numbers and their successors always exist, and that in turn might be taken as proof that recursion is not of necessity temporally dependent. In fact our purposes do not require of us a rigorous argument. It suffices to suggest that sequentiality and directionality sometimes function as an ersatz temporal element in such examples. The primary function of the introduction of time in (D2) and subsequent definitions above was to inject those two qualities, sequentiality and directionality, into the recursive context. To be useful, recursion in general requires a boundedness which is sometimes described as an "exit condition" by computer programmers. In addition, recursive procedures must be activated only at specific points in the course of a larger algorithm's "flow." A procedure which "called" itself either randomly or without end would be of little use in routines designed to solve real problems.

But sequentiality and directionality are ultimately very difficult to define. It is likewise problematic whether and how a pattern can be identified in such a way that its continuation can be accurately predicted. That is because sequence as pattern can be easily misapprehended: what we take to be the pattern displayed in a list of numbers may in fact be contradicted by the next element analyzed. Similarly, there may be any number of ways to define the apparent *telos* of a given series. If we have a heuristic notion of time, on the other hand, we can talk about sequentiality in relatively simple terms. What comes "next" (understood as a term which appeals to sequence) is what comes "after" (conceived as a reference to time); similarly, the final state of a sequence is what happens "last."

Do we lose anything by using terms which can describe temporal as well as sequential relationships -- *next* (*after*) and *last* -- rather than employing only specific, synchronic terms of pattern and sequence? The answer is simply that it depends on what the purposes of analysis are. A mathematician will in general not be satisfied by hearing that a given element in a sequence is followed by the *next* element nor that a finite sequence ends with its *last* element. But even if such statements are analytic, not all propositions employing generic language of this kind are uninteresting. In Wirth's examples above, as we will see in a moment, there is an interesting and vitally important linkage between general descriptors like *next*, *after* and *following* on the one hand, and the phenomenon of closure on the other.

The second and third of Wirth's examples likewise demonstrate the cumulative character of recursive objects consistent with definition (D3). We see a tree composed of smaller trees and a mathematical operation, factorial, as the repetition (accumulation) of binary operations using the previous value as the present argument of an unchanging function. It seems likely, then, that the first half of Wirth's definition of recursive object (D1.a) amounts to the definition given at (D3). Not only is (D3) a reasonable way of interpreting Wirth's mysterious "partially consisting" phrase, but it is also consistent with the first half of his own definition and with his examples of recursive objects.

Still, the notion of cumulativity in our latest definition of recursion might be called vague, since it does not account for our sense that an everyday composite object such as the average film is non-recursive. What do we mean by cumulative, or by saying (as in definition D3) that a recursive object consists partly of all its previous states? Obviously such cumulativity does not always refer to the literal additivity which we saw in Wirth's first example. It may help to turn to set theory, as we did when “unpacking” his notion of something “partially consisting of itself.” In that case, we used the basic notions of proper and improper subsets. In this instance, it seems that what Wirth’s three examples have in common is *closure* on a certain set with respect to a certain operation (Durbin 1979: 21). If a set is closed with respect to a particular operation, then whenever that operation is performed on members of the set, the result is another member of the same set. It is easy to see that the loci of Wirth’s examples are closed with respect to simple binary operations such as addition and multiplication. In the first example, the set in question is that of all natural numbers. Applying the operation which we might call “succeeding” yields another element of the same set. That could be said of Wirth’s second definition, as well, only here the operation is “joining”: join any two trees at a common node and the result is another tree. Finally, the third example boils down to the familiar fact that the set of non-negative integers is closed with respect to multiplication. Wirth does not analyze his own examples in this way, nor does he even mention the notion of closure, but it seems to apply.

It should be made clear that the definition of closure is all or nothing in theory, even though we often speak more loosely in practice. Under the assumptions of Darwinian evolution, for example, sexual reproduction viewed as a kind of binary operation on the members of a species is not necessarily closed. If it were, there would be no speciation. Under some conditions sexual reproduction is of course closed, but not under all. Geographical isolation of two breeding populations and genetic mutation -- what we might in general call “trump conditions” -- can ruin the closure. Let us define a trump condition as anything which (a) overrides expected closure in a single, identifiable iteration of an operation (e.g., genetic mutation) or (b) causes members of a set which was once closed on a certain operation such as sexual reproduction to produce an “element” which is not a member of the same set, but to do so in a gradual, almost imperceptible manner. The second type of trump condition challenges the law of excluded middle, at least in our perception, but does not violate it in fact. The key to seeing this is again the temporal aspect. Two recently separated populations of the same species will produce offspring which can reproduce with one another, while the offspring of two long-separated groups will be unable to do so (if we choose the proper durations for “recently separated” and “long-separated”). At some point in the temporal “middle” there will be an individual in the first population who cannot reproduce with a certain individual in the second population, even though their parents could have bred successfully. In any case, species are theoretically not closed with respect to sexual reproduction, even though we tend to think of them as being so for many purposes.

Binary operations on numbers are likewise not always closed, depending on how one defines "number." We saw in the introduction the Pythagoreans' difficulty with irrational numbers emerging from operations on natural numbers in certain contexts. There are also contemporary examples in which particular numerical operations yield results outside the range expected given the nature of the operands. An old joke among numerical analysts goes like this: "2 + 2 = 5 for sufficiently large values of 2." The phrase "sufficiently large" is a way of indicating that numbers which have been created or manipulated by computers are not always what they might seem. Even using features such as the FORTRAN language's "double-precision" means of defining real numbers, it is frequently impossible to include enough decimal places in the original definition of quantities to avoid horrendous round-off errors. If an algorithm is sufficiently repetitive, then a "number" can become much different than what we expect it to be.

There is, however, an obvious difference between this kind of round-off error and what we have called "trump conditions" -- a difference having to do with predictability. There is no *a priori* means of determining what will constitute trump conditions for, say, two populations of the same species. Observation and experiment may eventually provide after-the-fact data, but we cannot say in advance that separating two populations of the same species for X number of generations will yield two separate taxa (defined, following Mayr, as populations of animals which cannot or do not naturally interbreed; 1940: 254). Similarly, we cannot predict the precise occurrence of a mutation which will violate reproductive closure. If we find two species which we believe were once part of a single breeding population, we also cannot point to a specific moment and cause of divergence. Many round-off errors, on the other hand, are predictable and reducible simply because the environment (the particular computer) is so well known. If we were to employ a random number generator at some point in an algorithm and keep the sub-algorithm of that generator secret, then arguably unpredictable and irreducible round-off errors could occur. But otherwise, round-off errors can be theoretically if not practically predicted, or at least attributed to precise causes after the fact. This will become significant later; for the time being we return to our attempt to define recursion.

We can conclude from definition (D3) and from Wirth's examples that a recursive object is one which is built of its own previous stages and closed with respect to something. But to what? When we have stipulated in advance what the set in question consists of, then we also know how to understand the locus of cumulativity because we know what closure means for that set. In Wirth's examples above, we know that we are seeking closure on natural numbers, trees, and non-negative integers, respectively. The identity of the object in question is guaranteed by stipulating that the process of cumulative building is closed with respect to that object. In other words, the accumulation of "stages" will not yield a qualitatively different thing but will merely lend a "layered" character to the object itself.

We will develop the notion of recursive fitness at greater length below, but for now we should notice that if fitness can indeed be conceived as recursive, it will remain closed. There are no trump conditions which can cause an organism's fitness, conceived as a momentary status in an ongoing accumulation of relationships between phenotypes and environment, to become anything other than fitness values -- different ones, perhaps, but fitness values nonetheless. This is significant in that it provides a way of approaching what was called Meno's problem in the last chapter for a specific set of cases. Recall that the problem can be expressed as a question: How can we describe the exact relationship between a composite whole and its parts? The question was difficult or not impossible when it concerned the link between virtues and Virtue. But in cases where the relationship is recursive, the question becomes tractable in a certain way. A recursive object is simply itself, or rather slices of itself which accumulate over time in the way discussed above according to a certain governing function. But that raises the question of how we are to understand an object consisting of parts of itself temporally sliced according to an algorithm. In this context, the function which determines the timing and sequence of parts associated with the object amounts to a definition.

## 3. Self-definitional recursion

We have briefly examined the first half of Wirth's basic definition of recursion

(D1) -- what we have called "partially consisting recursion," following Wirth's wording, and labeled (D1.a) -- while noting that it is possible to view both objects and methods as recursive. The second half of Wirth's definition reads:

(D1.b) "An object is said to be *recursive*, if it ... is defined in terms of itself" (Wirth 1986.: 135).

What could be the difference between "partially consisting of itself" and "being defined in terms of itself"? As we have seen above, when we consider objects as being recursive in so far as they partially consist of themselves, it appears we must look at them as varying across time in such a way that they have one set of properties at one moment and a different set at a different instant. Moreover, the subsets which figure in partially consisting recursion are cumulative: a later instantiation of the object encompasses all of its earlier manifestations in some manner. This diachronically unfolding character is in some sense "real" rather than merely "perceived." The recursive object (so says the first half of Wirth's definition) *actually* consists of all earlier parts of itself in addition to whatever has been added to them to constitute the present manifestation.

The concept of self-definitional recursion will need to be developed at some length, but at this point it may help to offer a heuristic indication of how this kind of recursion functions. Recall that in the previous chapter we looked briefly at Aristotle's taxonomy of causation as a means of understanding multiple realizability. One of the questions we asked ourselves was whether fitness could function on either end of a relationship of multiple realizability. It seems clear, for instance, that we treat the same fitness value manifested in different organisms as an efficient causal explanation of why those organisms will fail or thrive in a given selective environment. Looking from the opposite perspective, we ask whether a given fitness value may be multiply realizable on a basis set of causes -- of material causes, for instance. What we witnessed, in short, was a circumstance in which one cause may be defined as a function of another cause.

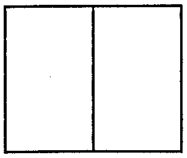
### (1) Defining as abstracting (perceiving) and creating

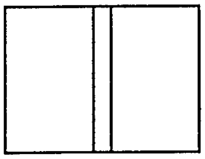
The phrase "being defined in terms of itself," by contrast, seems to entail the existence of a defining agent which can play two roles. First, the act of defining can be viewed as a way of carving reality at its joints, or at least trying to. The observer plays gardener and entomologist in the "blooming, buzzing confusion" of William James's well-known phrase, deciding what facets of experience to pluck or catch, and then how to describe the resulting specimens. We can call this aspect of definition *abstraction*. Secondly, the act of defining is tantamount to *creation*. The soup of experience already exists, but a particular distillate of it is something new, something made by constructing one particular definition instead of all the other possible ones. Definitions can be seen as ways of representing portions of reality, and how these representations are formed determines in what ways they are amenable to analysis. Some representations, for instance, are amenable to algorithmic solutions, while others may not be (Wagner 1994: 228 - 232).

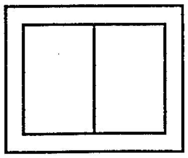
It is also apparent that a definitional understanding of recursion is epistemologically based in contrast to the ontological reality which we observed in partially-consisting recursion. That should not be a surprise. If (as this dissertation argues) recursion can function as a template for interrelating "layers" of experience and abstraction to one another within a locus such as that circumscribed by Darwinian fitness, we would expect to find both epistemological and ontological aspects in the general understanding of recursion. This is because many issues in evolutionary biology encompass a spectrum ranging from observational experience to inferential abstraction. If we wished to accommodate the experiential end of that spectrum a more purely ontological template would do; by the same token, as we penetrate deeper into the realm of abstraction a primarily inferential template will suffice. But to treat the entire spectrum as a unit, including the transition from experience to abstraction, we need a template which makes ontological as well as epistemological claims.

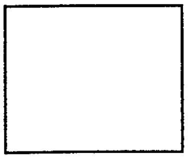
In the case of Wirth's definition at (D1.b), we see a special kind of definition -- one which describes an object in terms of itself. What would it mean for an object to be defined in terms of itself if we emphasize the act of defining as abstractive perceiving and creating? Recursive *perception* suggests the repetition of the same pattern of analysis on various, extracted "layers" of an object. The defining agent chooses which aspects of the object to single out, and the ones chosen are thus made somehow representative of the object as a whole. Again a temporal or serial element is critical, as in the case of recursive objects conceived as those which "consist partially of themelves." The same is true of definition as recursive *creation.* If defining a recursive object means bringing it into existence by identifying the sense in which it consists of self-representations, then we can easily distinguish at least three rough periods -- before the object exists, the period of its creation, and the time after it exists. Furthermore, each of these periods may be seen as having a duration. The middle period, for instance -- the process of creating by defining -- is itself a temporally- or serially-bound activity.

An illustration lies ready to hand. Presumably this manuscript lies open on a table or lap. (There is no need to extend the example to encompass other possible forms, such as a projection on a CRT, but we could easily do so.) The manuscript's overall form can be perceived as two rectangles (sheets of paper) joined in the middle, as in the first figure below; or as the bigger rectangle, showing two sheets joined by a "binding" rectangle lying between their inner edges, represented in figure 2; or as a still bigger rectangle formed by the edges of the dissertation's cover surrounding the two-sheet configuration described in the first figure (figure 3). Those could all be described as recursive ways of viewing the manuscript's overall form, and of course we could imagine any number of additional ways of perceiving the manuscript's two-dimensional appearance as a sum of rectangles. We could choose to see the margins as rectangles, for instance, or we could view the printed portion of each page as a rectangle. We could apprehend each line of printing or even each letter as a rectangle, along with any combination of these possibilities. Something like figure 4, by contrast, represents a non-recursive perception of the manuscript's form. It merely shows the outline of the manuscript without suggesting that the shape is a composite of like forms.


1


2


3


4

Of course this is merely a metaphorical explanation of what constitutes recursive definition as perception. Not only are the figures above an obvious oversimplification of what the reader of this dissertation sees on the table before him, but they also represent a model of perception rather than perception itself. A neurologist might offer a much different and more detailed model, but let us accept at least the possibility of recursive perception as defined above -- the act of apprehending a thing as a composite of units which are like itself in some significant way.

Adequately exemplifying recursive definition as *creation* seems inevitably to presuppose recursive *perception.* An illustration may help make this clearer. Perhaps a member of a forest-dwelling tribe in the Philippines forty years ago had never been introduced to the concept of a rectangle. Assume further that the tribesman in question had never seen a near-perfect rectangle nor had he imagined a perfect one. A missionary arrives one day and begins teaching the obligatory curriculum of religion and Spanish. In the course of the lectures, however, the curious tribesman inquires about the appearance of houses and churches in the world beyond the forest. An *ad hoc* geometry lesson begins, during which a number of figures are scratched out in the dirt of the village "square." Among the figures demonstrating the appearances of houses and churches is a rectangular wall, with a rectangular door, rectangular windows, covered by a rhomboid roof (can't win 'em all) which is topped in turn with a rectangular chimney. At least the wall of the house, if not the house as a whole, can be thereby defined -- and *created* as a first-time image for the tribesmen -- as a rectangle containing other rectangles. The facade of a skyscraper, the overhead view of a modern city, and any number of other new images might be similarly defined/created for the tribe's instruction.

What is not quite clear in all of this is how we are to distinguish between self-*reference* and self-*definition.* This time an example lies closer to hand than the Philippines. The table of contents (pp. 2 ff. of this manuscript) refers to itself. That is, the part of the dissertation responsible for indicating where each part of the dissertation is to be found offers us its own location. Is the table of contents therefore a recursive object? Intuitively, the answer is no. The table mentions itself but does not fulfill the criterion necessary to be recursive in a partially-consisting way since it is not a diachronically unfolding accumulation of its earlier manifestations. Similarly, it would be difficult to define the table of contents in terms of itself. The table seems not to exemplify definitional recursion as abstractive perception, since we do not apprehend the table as a diachronic accumulation of things like itself. Nor is the table definitionally recursive in a creative sense, since we cannot imagine an engineering process which would create the table from references to itself. We can conceive of the table as a composite of sub-tables (of sub-tables of contents for each of the three main parts of the dissertation, for instance), but that is always possible, even where no self-reference is present. The question then presents itself: If definitional recursion is not manifested each time a *definiens* merely makes mention of its *definiendum*, what additional conditions must exist?

### (2) A causal factor in definitional recursion

Recursive representations, whether verbal or visual, unfold serially. That is simply to say that the act of creating by recursively defining must be performed diachronically. This seems to imply that the definition we constructed above to explicitly address the temporal element (D2) still applies, though perhaps with the italicized addition in the following definition:

(D4) An object O: $\{o_1, o_2, \ldots\}$ can be recursive if at any time $t_i$, O consists of all its previous subsets. *Moreover there must be a subset of O:* $\{o_k, o_{k+y}, \ldots\}$ *which causes the existence of the same or another subset of O at some* $t_m$, *unless an exit condition has intervened.*

In other words, there must be a causal link between one iteration of the recursive algorithm and the next. That is not to say that one iteration inevitably follows on the heels of another; that would be tantamount to what programmers call an infinite loop, meaning a procedure with no exit condition. Rather, I want to claim that *if* an iteration exists, then it was caused by a predecessor iteration which had the same form. In order to make this concrete we can return to the examples employing rectangles. To the extent that such examples instantiate definitional recursion through perception or creation, they do so not merely by implying that a given stage of the rectangular structure would not have existed unless the previous stages had also existed. That only repeats a necessity condition which we have already discussed -- namely, that a recursive object is an accumulation of its predecessor states. The causal element arises when we stipulate a sufficiency condition as well: Whenever the algorithm has reached a certain stage, then it will inevitably proceed to another iteration. This contention requires justification.

It seems clear that if something is defined in terms of itself in the cumulative fashion discussed above, then at any point in time one of two assertions must be true: (a) either the object has yet to undergo its first recursive iteration, in which case it has no constituent parts in the recursive sense; or (b) the current state of the recursively composite object implies all the former states necessary to have constituted the object as it exists. To illustrate this causal relation, suppose that a certain school of pedagogy emphasizes repetition of previous content as the most efficient method of learning. A given curriculum for a 5-month course might look like this:

<u>Monday</u>

In class: instructor reads the first page of Kant's first critique aloud.

<u>Tuesday</u>

Same as Monday, then read the second page.

<u>Wednesday</u>

Same as Tuesday, then read the third page.

<u>Day X</u>

Same as day X - 1, then read the *next* page.

<u>etc.</u>

Notice that this plan does not say that on day X the student should read the first X pages. Rather, the student would read page one, then pages one and two, then pages one through three, then pages one through four, etc. Thus the reading assignment on day *n* as well as the content of the curriculum as a whole might be described as a function of the term $\sum_{i=1}^{n}$ $p_i$. In some sense all previous readings, representable in terms of the expression $\sum_{i=1}^{n-1}$ $p_i$ are the "cause" of the curriculum on day $p_n$. Thus all that the cumulative character of the recursive algorithm requires can be expressed as $\sum_{i=1}^{n} p_i \Rightarrow \sum_{i=1}^{n-1} p_{i.}$ But the causal criterion in (D4) goes further and claims that $\sum_{i=1}^{n} p_i \Leftrightarrow \sum_{i=1}^{n-1} p_i$ unless an exit condition interferes. In other words, given the definition of the curriculum described above, every previous state is a *sine qua non* of the present one and is therefore a cause of the succeeding state. In fact, a previous state in a recursive definition is a particularly robust kind of cause -- what we might see as being at once material (though perhaps in a metaphorical sense) and efficient (to employ Aristotelian language) in relation to the present state. The present state, in turn, may be seen as formally (isomorphically) the same as the previous state and as final (fulfilling the role of *that for the sake of which* the previous states exist).

## 4. The "threshold" of recursion: object-and-method, value-argument-function

It remains to discuss an issue which we have glossed over thus far. In the essentially visual examples involving the two-dimensional appearance of a manuscript and the wall of a house above, it was not clear that we observed something which had been defined in terms of itself. Rather, we merely understood some property of the constituent objects as representing the *essence* of the *definiendum* while ignoring other qualities. For instance, we took the manuscript as a rectangle to consist of other, smaller rectangles. The shared "rectangle-ness" of the object and its parts engaged our full attention; we did not consider the obvious differences in size and proportion as violating the strong criterion of identity or the weaker criterion of similarity implied by the key phrase in the second half of Wirth's description of a recursive object -- being "defined *in terms of itself*." It is not clear that the rectangle example as presented above fulfills even the weaker criterion of similarity between the object and its constituents. The same thing could be said of Wirth's own example of "an advertising picture which contains itself" (1986: 135) and of Hofstadter's Russian dolls (1979: 127). All that can be said with certainty is that the strong criterion of identity has not been met; like the rectangles, the progressions of embedded images and dolls comprise units which are different from one another despite their similarities.

On what grounds do we make the leap from conceiving an object as merely composed of parts somewhat similar to each other to seeing a composite thing as consisting of or defined "in terms of itself"? No one of Hofstadter's Russian dolls seems to be the sum total of all the dolls; no embedded picture in Wirth's example is identical to the collection of all pictures. One resolution is to stop demanding identity of objects and insist instead that a function (algorithm) be given which dictates the recursive link between objects; the function then becomes the shared attribute which suffices to meet the criterion of self-definition. Such a function cannot stand alone, of course. By definition its identity as manipulator and yielder of quiddities must encompass objects. In particular, a function must be linked to an argument (or arguments) and a value. It is of the utmost importance here to recognize Frege's insight (1891): A function is *not*, strictly speaking, the same as its value. By Frege's account, treating the two as the same is merely a convenience, and often a confusing one.

We will try to avoid this potential confusion by bearing in mind that we must distinguish between object and method while treating recursive scenarios as a combination of both. Object, moreover, refers to two distinct aspects of quiddity -- value and argument. A recursive context must therefore be described in terms of three aspects -- algorithm, value, and argument -- and in such a way that the argument at any moment (iteration) is the value of a previous moment.

To see this more clearly, consider other metaphorical examples of recursive objects. Frequently these simply show repetition of what is essentially the same pattern. M. C. Escher, who appears in the title of Hofstadter's 1979 book and whose work graces many of the book's pages, produced a number of works based on such repetition -- for instance, a finite number of birds which, when placed next to each other, rotated appropriately, and repeated, fills the entire picture without the need for what an artist would term "negative space."[15] (Artists use this term to describe what a picture contains in addition to its primary image or images. Sometimes, but not always, negative space is synonymous with background.) Or to describe the same phenomenon differently, the repetition of bird images in Escher's drawing makes the distinction of positive from negative space a matter of perceptual choice.

It could be said that the whole of such a picture is defined in terms of a repeated pattern, but that is a far cry from the claim that the picture as a whole is defined in terms of itself. Not all aspects of the picture may be given by defining it in terms of any subset of itself, unless of the *improper* subset, that is, itself. But saying that the picture is itself, A is A, would be no explanatory coup. Alternatively we might choose to define the picture as a repeated pattern. That kind of definition can be framed in words indicating that the *definiendum* and the *definiens* are *similar*, but a strong contention of *identity* is indefensible. "Stepping this way, ladies and gentlemen," a tour guide in a museum might announce, "you will see a drawing done by M. C. Escher in 1942. It is a repetition of a finite number of birds, variously rotated." That is not a recursive definition even in a weak sense.

In order to formulate a recursive definition, the guide would have to claim something such as: "Here is a thing having to do with birds composed of things having to do with birds." (Here is an A composed of As.) That tactic is unsatisfying because it *could* be applied in a loose way to many pictures which, intuitively, do not exemplify recursion. For instance, we could look at a Seurat (or any pointillist work) and say, "Wow! What a great visual metaphor for recursion: an overall picture conceived as a big dot of paint composed entirely of little dots of paint!" Again intuitively, we would prefer that the term recursion be applied to a more restricted range of visual metaphors, else the purpose of providing such visual examples -- to understand what recursion is in the abstract -- will go unfulfilled.

Unfortunately, examples such as those involving rectangles and Russian dolls suggest that if we have correctly interpreted Wirth's basic definition, it is very difficult to find an indisputably recursive object. For if the object at any given time is an accumulation of its previous states, each manifestation of the object must be new and different from earlier manifestations in some way. Hofstadter's Russian dolls grow increasingly larger, else the "later" dolls could not accommodate the "earlier" ones. The same is true of the rectangles which constitute this manuscript's two-dimensional shape: they differ not just in size, but in their proportions, too. This is a worrisome business. If we treat Russian dolls and rectangular assemblages of rectangles as recursive objects, but reject one of Escher's pictures consisting of a repeated pattern, do we base that taxonomic decision on a rigorous criterion or on a whim? At bottom such a decision seems to be based on an intuition which can be codified as a further refinement of Wirth's basic definition:

(D5) A necessary condition for an object to be judged recursive is that it and all of its constituent parts must be describable in essentially the same way, that is, as sharing the same quality or qualities.

Alas, this statement is filled with loopholes, the largest being the qualifier "essentially." The heuristic motivation for being suspicious of any attempt to treat (D5) as *sufficient*, too, has already been discussed, but it won't hurt to repeat it in slightly different terms. Suppose there is an organism or machine (it does not matter which) that "perceives" its world by answering a Boolean question of the form

(Q1) "Do objects X and Y have the same _____?"

The blank can be filled in with any quality in the perceptual domain of the perceiving organism or machine. Taking the simplest case first, if the order of the perceptual domain is one -- say the perceiver "sees" its world only in terms of size -- then it is apparently impossible for a *properly* recursive object to be perceived. (Recall the definition, given above, of a proper subset: one which is part but not the whole of its parent set.) A perceiving agent having a perceptual domain of one category (from now on let us call such a being an order-1 perceiver) would survey Hofstadter's Russian dolls and return a clear verdict. If the order-1 perceiver's criterion of judgment is size, the dolls fail to fulfill the standard set forth in (D5) and (Q1) since they are all different from one another in the only aspect perceived. (We can take it that the two, definition (D5) and question (Q1), boil down to essentially the same test and may therefore be called a single criterion.) For such a perceiver the "dolls" do not really exist as dolls. There is simply a set of quantities ranging from small to large: {littlest, littler, little, bigger than little, ..., littler than big, big, bigger, biggest}. Since none of the dolls is the same as any of the others with respect to size, there is no chance of perceiving a shared attribute aside from the category of perception itself (all elements exist in 3-space, so all have size).

If the order-1 perceiver used proportion (congruency) or coloration as its perceptual standard, on the other hand, we can imagine a set of Russian dolls which would meet the sole requirement of (D5) and (Q1) -- sameness. In this case, however, an order-one perceiver would hold the dolls to be *entirely* the same. The *set* of dolls, by contrast, is nothing more than a collection of identical elements, but no element is the same as (nor perhaps even similar to) the set.

Now let us try out an order-2 perceiver, that is, one which is sensitive to two attributes -- say, size and color. If we assume that each doll in a set is like every other doll in both respects, then the case is the same as the one involving an order-1 perceiver. If, on the other hand, Hofstadter's Russian dolls are precisely the same color even though they differ in size, perceiving the dolls in terms of their shared feature (color) can *make* them fulfill (D5) and (Q1). If this happens, we might say that epistemology overrides ontology. But the order-2 perceiver has the ability to notice, as well, that the dolls differ with respect to size. From this perspective, the collection of dolls still cannot constitute a recursive object, but that is not solely because the order-2 perceiver will see each element of the composite whole as similar to but not identical with the other elements. (In that regard the order-1 and order-2 perceivers are similar.) Rather, the non-recursive character emerges because no object is the same as the collective case. Here the relationship between the parts of the composite whole is always one of identity or similarity, but not of recursion.

This raises a question which expresses the dual challenge of understanding how an object can partially consist of itself (section 1 above) and how it can be defined in terms of itself (section 2): How can objects stand in a recursive relationship to one another? The answer depends on treating the relationship as an epistemological overlay on the objects understood as existing ontologically. It could be said that the observer forces a recursive character on a reality which could also be perceived non-recursively. (This will strike most computer programmers as being wholly obvious. They know from experience as well as from an understanding of theory that a recursive algorithm can always be rewritten in iterative form, and vice versa. The choice of which way to implement a series of operations has to do with ease of understanding and aesthetics. Even though there are rules of thumb -- e.g., it is especially easy to rewrite "tail-recursive" algorithms, in which "the recursive call is the last statement in the procedure," as iterative algorithms (Bentley 1988: 31, 163) -- the decision as to what constitutes "elegant" as opposed to "awkward" implementation is more a matter of art than science. Likewise, perceiving a relationship as recursive is a choice made in an epistemological context; such a way of perceiving is not forced by the observer-independent ontology of the situation.)

All of this is consistent with the discussion of definitional recursion (section 2 above), in which it was suggested that recursion is never wholly ontological nor wholly epistemological. Further, we have seen that Wirth applies the qualifier "recursive" to both objects and methods. (Unfortunately he does this with abandon, scarcely seeming conscious of any tension between the two perspectives and without analyzing the difference.) By breaking down the category of "object" into value and argument, and understanding method as algorithm, we have the elements necessary to understand fairly precisely what recursion means and how it can apply to loci of data such as those used to determine the fitness of organisms. But before approaching fitness itself as a recursive object-plus-method (henceforth it will be understood that fitness has this composite character regardless of what phrase is used -- e.g., "recursive thing," "recursive algorithm," "recursive object"), we will remain at a more or less abstract, general level to inquire how the concepts of reduction and supervenience apply to recursive contexts.

# Chapter Four: Reducibility and Supervenience in a Recursive Context

## 1. First Definitions of Reducibility and Supervenience

The substance of this chapter is probably somewhat predictable given what was said in the introduction together with the analysis of multiple realizability and recursion provided thus far. In a nutshell the argument here will be that recursive contexts are not necessarily reducible in any *interesting* fashion beyond the reduction offered by the recursive scenario itself. A recursive triad (function, set of values, set of arguments) already *is* a reduction in some sense: it is a full *formal* account of how the quiddities involved -- the values and the arguments -- are related to one another by a certain algorithm. But there is nothing ontologically deterministic in the relationship between the quiddities, since the function -- the epistemological choice of definition -- controls that relationship. On its own formal level, the function itself provides a predictable reduction. It says that the "thing" in question is a function of itself. The particulars of such a situation may be interesting nonetheless, but that character depends upon the purposes and predilections of the particular researcher involved. For instance, someone whose first introduction to a Fibonacci sequence is through a recursive function might find the specific form of that algorithm fascinating. But as we will discuss below, the presence of such a function does not entail the possibility of a complete *and interesting* reduction.

### (1) Complete and Interesting Reductions

An example may help make clear what *complete* and *interesting* mean in this case. Wirth (1986: 153) offers the following "crude version of a solution" to the "eight queens problem" (How can eight queens be placed on a chess board such that none checks another?) which Gauss attempted to solve in 1850:

```
PROCEDURE Try(i: INTEGER);
BEGIN
  initialize selection of positions for i-th queen:
  REPEAT make next selection
    IF safe THEN SetQueen;
      IF i < 8 THEN Try(i + 1);
        IF not successful THEN RemoveQueen END
      END
    END
  UNTIL successful OR no more positions
END Try
```

In essence this procedure says that to find a solution we should place the first queen somewhere, verify that she's unchecked or put her someplace else if she's checked (the latter possibility can't occur for the first queen, of course, but can for all the following ones), then move to the next queen and repeat the procedure until we run out of safe squares or queens. But the algorithm does not tell us everything there is to know about initial conditions. For instance, suppose we look at a board arrayed as follows (ibid.: 154):

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Q | | | | | | | |
| | | | | | | Q | |
| | | | | Q | | | |
| | | | | | | | Q |
| | Q | | | | | | |
| | | | Q | | | | |
| | | | | | Q | | |
| | | Q | | | | | |

This is clearly an object-solution to the problem while the procedure above is a method-solution, but even given this object and the recursive algorithm which led to it we cannot say which of the queens was placed first, in case that is of interest to us. Knowing how programmers think and work, we might guess that the first queen was placed on a corner. But in so thinking we *might* err. The only means of being certain would be to know the initial conditions under an eyewitness codicil.

It should be noticed that one might treat Wirth's "crude" solution to the eight queen's problem as a measure of fitness on a 64-square environment which exerts selective forces: checking, threatening, capturing (killing), restricting movements (barring the queen from moving as a knight would, for instance). We can choose to understand the fitness of the species of queens here -- call them *Wirthi reginae* -- in a Boolean way. Either they all make it onto the board or not. Or we could throw a queen-counter into the algorithm (initialize the variable *QueensPresent* to 0, then add one each time a queen is successfully placed) and thereby gain the ability to read off a numerical expression of the species' fitness: if only six queens can arise in an environmental context imposing harsher selective forces (perhaps a board with fewer squares in this case), then the fitness of the species in general is less in that environment than on the 64-square board. Similarly for the fitness of individuals: a Johanna-come-lately queen (an individual who tries to settle in the environment anytime after the first eight individuals already are established) will be decidedly unfit on the 64-square board.

There are any number of other ways in which we might choose to measure the fitness of queens as individuals or as a species. For instance, we could easily write a space-counter into the algorithm (initialize *SpacesTried* to 0, and increment by one every time a space is tried but rejected). With such a tool at our disposal we might elect to measure an individual's or the entire species' fitness by finding out how many squares had to be checked before the queen or queens in question were safely settled. The list of aspects we could measure by modifying the basic algorithm goes on and on. Moreover, the information thus yielded mimics data which are of real interest to practicing evolutionary biologists, who want to know all about the amount of time and energy organisms expend to move into and survive in a given selective environment.

We should remind ourselves, however, of what was said above about the reducibility of the solution provided by Wirth. What goes for that specific example holds in general as well: knowing a state of affairs and having an idea about the algorithm which achieved it do not automatically allow us to reason backward and uncover every datum which might interest us. In a large computer program, it may well be that some variables are initialized outside of a particular module (algorithm).

Even if that module is no black box -- i.e., even though we have insight into its mechanism -- and even if we have its output for a certain period and know the kinds of things it must have manipulated, we still may not be able to reason backward to the initial state of the variables it controls. In other words, we cannot necessarily achieve a complete reduction in the sense of one which transcends what is already offered in the formal function-value-argument triad. Furthermore, the possible (formal) reduction may not be of interest. To make this more concrete, we can refer to the statement of a praciticing ethologist (whose term "constraint" is analogous to what we have called "initial value").

> "I...wish to emphasize one principle which is explicit in [Grafen's] analysis as well as implicit in that of Trivers and Hare. The question is not 'Has the "best" sex ratio been successfully achieved?' On the contrary, we make a working assumption that natural selection has produced a result, given some constraints, and then ask what those constraints are...." (Dawkins 1982: 76)

Notice that the constraints cannot be known with certainty, that is, we cannot uncover them through reduction. At best we can infer possibilities consistent with extant data relevant to the case under consideration and with whatever models are applicable. In the case Dawkins reports, for instance, there is a principle, natural selection, which is taken as determining a large set of behaviors and structural features of organisms. The principle is simply assumed to be operative when observing a particular behavior. If that behavior seems not to conform to a simple interpretation of the principle, one does not question the principle itself, but rather seeks to find "constraints," that is, factors which will explain that the principle still works, albeit in unexpected ways. But there is no guarantee that all constraints can be uncovered. Some, namely those which are one-time occurrences (analogous to the initial conditions which led to a particular placement of queens) may well be unrecoverable.

These reflections yield senses of both complete and partial reductions. Given a scenario and a function which controls its dynamic, if we can reason backward to answer *any* question which interests us, then the context is *completely* reducible. But if some questions which are of conceivable interest cannot be answered with certainty, then the context is only *partially* reducible. Notice that this distinction between two kinds of reducibility does not have to do with temporal proximity or complexity. If a super-computer solved the eight-queens problem in an almost vanishingly small instant of time, there still may be questions which we cannot answer even though we have the computer's solution algorithm and Wirth's placement of queens. On what square was the first queen placed? How many queens have tried to enter the environment but failed? The data necessary to answer those questions are not to be found in either the algorithm or the solution, even though the algorithm produced that solution almost instantaneously. Nor are all incompletely reducible scenarios intractable because of the complexity of the problem relative to the cognitive abilities of the observer. The unanswerable questions mentioned above would still be unanswerable if we confronted the much simpler problem of placing five queens on a 25-square board, unless the initializations happened to occur where we could see them (the eyewitness codicil again). This locus of distinctions between completely and incompletely reducible contexts thus differs from the apparent grounds of, say, Weber's (1996) distinction between microreductions and other, broader reductions.

## (2) A first look at supervenience

Just as the degree of possible reduction depends on what is given in the algorithm associated with a scenario and information about the algorithm's arguments, a recursive context in general exemplifies a relationship of supervenience rather than one of empirical identity in so far as arguments and values are not related apart from the recursive function which they complement. In other words, in a recursive context there is no possibility of one thing supervening upon other objects in the absence of the reductive function. One way of looking at this dependence is to conceive of the supervenience relationship as a kind of knowledge. If such a relationship were to arise independently of the function conceived as a means of solving something like the eight queens problem, what is sometimes called a "knowledge paradox" would exist. "Knowledge paradoxes violate the principle that knowledge can come into existence only as a result of problem-solving processes, such as biological evolution or human thought" (Deutsch and Lockwood 1994: 71). But what a basic *recursive* function tells us is simply that at any given moment a thing's value is related to earlier instantiations of itself. The function will not necessarily tell us all the particulars of a given relationship once that state of affairs already exists. Still, we can always say that in the context of the function-argument-value triad, a value is formally supervenient on preceding values.

The two definitions of supervenience which will emerge from the coming discussion are simple, perhaps deceptively so:

(1) A supervenience relation is any non-identity relation.

(2) A multiply realizable supervenience relation is any supervenience relation which conforms to the first chapter's conclusions about multiple realizability.

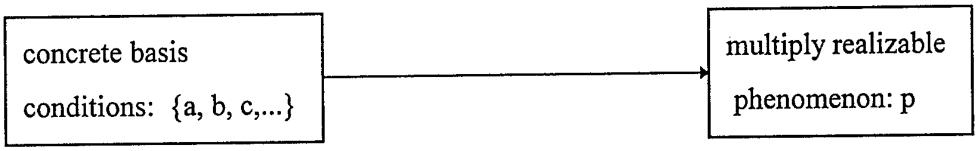
Definition (1) coincides with Weber's remark that supervenience relationships can be one-to-one, that is, that they need not be multiply realizable (1996: 417). In this I believe that he, in turn, is consistent with Frege's careful distinction between function, value, and argument (1891). What is interesting about supervenience in a recursive context is that it is impossible to say what the ultimate "foundation" of the supervenience relationship is. Certainly a status quo such as that represented by Wirth's solution to the eight queens problem is in some sense supervenient upon initial conditions, but the nature of the recursive algorithm makes any particular relationship to specific initial conditions more or less *accidental*. The algorithm would have worked no matter how we had initialized the variables involved, or at least for a wide range of initializations. Thus it would seem that the relationship between any value-argument pair is somehow more closely tied to the algorithm than to the initial conditions. Granted that the solution may have looked a bit different if the first square tried were (6,6) rather than (1,1), nevertheless the algorithm still bears an intuitive closeness to the solution obtained. Moreover, every iteration of the algorithm supervenes in some sense on every other: there is a dependency between every pair of iterations and the relationship is not one of identity.

If we measure the fitness of the *Wirthi reginae* or of any particular queen in one of the ways suggested above, then it appears that these numerical expressions of fitness are inevitably supervenient upon any number of previous states, ranging from the initial conditions (in case we happen to know them under an eyewitness codicil) through the iteration immediately prior to production of the fitness value in question (e.g., through incrementing a counter variable). However, these supervenience relations cannot be described independently of the recursive function which linked the units of relationship to one another. Those units are the various arguments and values of the function.

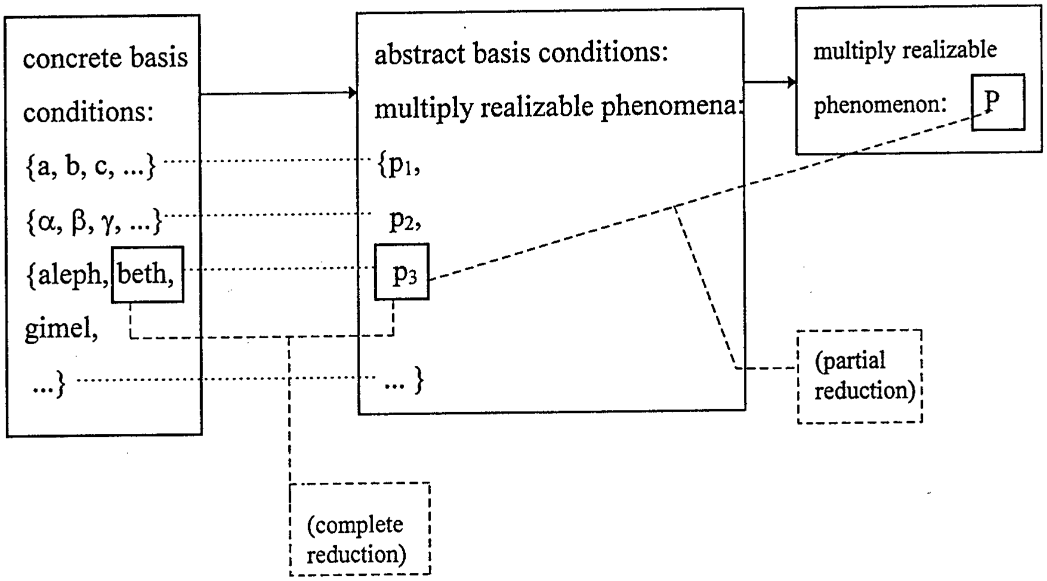
The same could be said of many or perhaps all functions, of course, not just recursive ones. What is unique about recursively supervenient relationships is that they entail iteration. It would violate the dictates of common sense and parsimony to say that something is a function of itself if the function in question were simple equality. Rather, as we saw in the previous chapter, a recursive function relates a value to previous but non-identical instantiations of itself, which means that at least one iteration of the recursive function must be made in order to generate the difference between the value and the argument. The resulting layered character, in turn, allows us to emphasize that a supervenience relation of a scenario to its initial conditions is always mediated. As we will discuss at greater length in an upcoming chapter, that self-mediating character entails a kind of circularity, but not a vacuously analytic or tautologous kind.

## 2. Reducibility as the other side of the multiple realizability coin

We may take the phenomenon of multiple realizability as a many-to-one relationship between a basis of realizing conditions and a phenomenon. The link may be what we have called direct or layered. In either case multiple realizability exists because more than one way of slicing and dicing the basis set will yield the same phenomenon. If we have a *particular* instance of that phenomenon, however, then it is natural to assume that of all the possible realizing conditions, only one is *actually* responsible for the phenomenon's existence *this time*. At another time, in another context, the phenomenon may have a different basis. In the case of Darwin's exchange with Gray over the Umbelliferae, for instance, a specific ratio yielded the figure two percent; in a different context, an uncountable number of ratios could have yielded that same percentage. Thus in particular instances -- those where an eyewitness codicil applies -- we can go backward (so to speak) and conceptually reduce a multiply realizable phenomenon to a specific cause or set of basis conditions. But where we do not happen to have special knowledge, we cannot always reason backward all the way to initial conditions. Or we might wish to speak even more generally and literally, so that reduction will refer to the *limiting of causal possibilities*. In this way we accommodate even cases in which we may not be able to discard all but one of the alternatives. We might call such reductions, in which we can narrow the field somewhat but not all the way down to a single basis, "weak reductions." (Using the example from chapter one, for instance, I might say that the sensation I just felt was caused by either the cat or the wind, but I rule out dogs and neural disturbances.) It is also conceivable that in a context involving what we have called layering, reduction might be possible to a certain point, but not all the way down to a single subset of the basis set ("broken-elevator reduction").



reduction in the context of
direct multiple realizability



reduction in the context of
layered multiple realizability

If we consider the context of multiple realizability as being metaphorically two-dimensional (as in the diagrams labeled "direct multiple realizability" and "layered multiple realizability" above), then we can at least attempt to analyze a multiply realizable phenomenon by moving in the opposite direction than that in which the arrows point. In other words, instead of a perspective which sees some single variation of the basis set as yielding the multiply realizable phenomenon through a certain function (e.g., in a truly disjunctive case), we can look in the opposite direction -- from multiply realizable phenomenon to basis set -- and ask what function if any can control the "reversal" from a multiply realizable phenomenon back to a specific subset of the basis. But if we do not know the specific initial conditions which led to a given solution to the problem, nor those which could have provided the same solution but in this case did not, then it is not clear what reversal means. For example, Wirth's procedure for solving the eight queens problem suggests its own reversal in more than one sense. Instead of searching for a way to place eight unchallenged queens on the board, we could "reverse" the part of the algorithm which guards against challenged placement and move the queens into a mutually challenging position. Or we could "reverse" by searching for and removing queens. The point is that we cannot equate reduction with the construction of a reversing algorithm.

Apart from the question of whether the phenomenon of multiple realizability is a boon or a headache for the student of nature, we can see that it would be misleading to claim that Darwin's chief concern in his remark to Gray was multiple realizability *per se*. Rather, what concerned Darwin was the *irreducibility* of some multiply realizable phenomena -- percentages, in this instance. As we have seen, there is no way, no algorithm, which allows us to go backward from a claim such as "Two percent of all indigenous plants in area A belong to the Umbelliferae" to a specific proposition of the form "In area A there are X kinds of indigenous plants, of which Y belong to the Umbelliferae." Any number of values for X and Y would make the proposition true if the sole standard of veracity were a claim asserting that a specific percentage applied. Thus we should distinguish between two kinds of multiple realizability. Some of the examples of multiple realizability which we have considered *are* reducible, either by appealing to the eyewitness codicil (I have privileged knowledge of the cause of the sensation I just felt -- It was that cat!) or because of the nature of the multiply realizable context itself. For instance, we assume that there is a truth of the matter with respect to the ancestry of birds. Given enough clues and a sufficient theoretical background, we can rule out some potential ancestries, perhaps even to the extent of drawing a single, very detailed path. In short, one class of multiply-realizable phenomena is reducible while the other is not. Darwin was undoubtedly correct in claiming that he needed something more than percentage values to pursue the research program he had envisioned. But the contemporary entomologist Douglas is just as convinced that although various functional utilities could have led to the development of modern insect wings, thermoregulatory effect is the single correct answer to the riddle. By mentioning the matter to Gray, Darwin tried to head-off future cases of irreducible multiple realizability (those in which no eyewitness codicil would provide the naturalist with the raw numbers corresponding to a percentage value), while Douglas believes he has reduced the range of possible causal explanations to a single right answer. In either case, what matters to the practicing biologist is not so much that multiple realizability exists, but rather whether the multiply realizable phenomenon in question is reducible or not.

## (1) The realms of reducibility

But here caution is required to distinguish between what we might call epistemological versus metaphysical reducibility. It is vague to claim, for instance, that corresponding to each *reducible* multiply realizable phenomenon P there is a set of sufficient but unnecessary causes C (consisting of $c_1$, $c_2$, ..., $c_n$) and a function or set of functions F which links P to a *specific* $c_i$ in a certain instance. (In another case, P may be "reducible" to a different specific cause.) In some cases, the reason why we cannot completely reconstruct the past is essentially epistemological, as Pääbo implies when he appraises a "Jurassic Park" scenario in which dinosaurs are reconstructed from excavated bits of DNA:

> It is my firm conviction that such dreams (or nightmares) will never be realized. We have no idea how to piece together the millions of DNA fragments that we extract from an animal into chromosomes in a functional cell, nor can we set in motion the thousands of genes that regulate development. If we cannot even take a fresh cell

> from an adult vertebrate and use it to clone another individual, how can we imagine cloning an extinct species from the flotsam and jetsam of ancient DNA? (1993: 92).

Clearly the emphasis here is on epistemological irreducibility. There is presumably a way to reconstruct the past from extant DNA corresponding to a fact of the matter about how the organisms *were* actually put together, but we do not know how to do it because we do not understand enough about genetics in general.

Similarly, murder mysteries and horror stories often begin with the illusion of irreducible multiply realizable instantiations. Something disturbing -- a fresh cadaver is an old favorite -- turns up in a closed system such as the island in Agatha Christie's *And Then There Were None* (1939) or the snowbound mansion in Stephen King's *The Shining* (1982). Since the system is closed, all of the dr*amatis personae* remain present, ergo someone in the frame of reference must be the perpetrator. Even more intimidating are some of A. Conan Doyle's tales, in which virtually anyone in all of London or on a vast moor or wherever a particular plot plays out might be the miscreant we seek. The epistemological challenge facing the reader and the tale's protagonist is to prove that in fact the murder or robbery or what have you can be linked (reduced) to a single causal agent. However, the reader knows from the first page on that there is no *ontological* challenge. In such cases there is always a fact of the matter, which is to say that the murder is always metaphysically reducible: although anyone present might have committed the murder *so far as the reader knows*, only one person did (at least in the classic formula for murder mysteries).

It is doubtful that this genre, sometimes referred to as "whodunits," would have any appeal if the writer's goal were simply to demonstrate that it is epistemologically *impossible* to figure out who in fact "dunit." Imagine that after three hundred pages Holmes takes a drag on his pipe, fixes Watson with a steely gaze, and says, "My dear fellow, it is elementary that the clues available are simply insufficient to solve the case. So let's call it a day, leave the whole mess to Lestrade, and go get a couple of pints." Such is not the stuff of bestseller lists. What readers long for is to see Holmes' sharp wit overcome what to any other mortal would be the apparent epistemological irreducibility of a multiply realizable incident. It would be even worse, from a marketing point of view, if the reader reached the end of the book only to discover that the mystery was in fact *ontological* rather than epistemological. It is hard even to imagine what such a story would look like in the realm of human affairs. What would it mean to produce a murder victim and simultaneously to claim that no definite causal agent committed the crime? The same sort of question seems to bother Rosenberg (1985, 1994) when he considers the problem of whether propositions in biology can be reduced to claims in physics and chemistry and Weber (1996) when he treats the possible microreducibility of fitness values to concrete circumstances: What would a proposition in biology be about if not, at its most basic, the same sorts of things as statements in physics and chemistry are about? And if biology, chemistry, and physics all share the same domain, then one may well be reducible to one or a combination of the others, or so it may be thought. A microreductionist such as Weber reasons similarly: Where would the fitness of an organism come from, if not from causes identifiable in terms of concrete phenotypes and environmental forces?

To clarify and summarize, let us accept that there are two different types of multiple realizability with respect to reduction -- there are multiply realizable scenarios which are reducible and others which are not. In this second class, that of irreducible multiply realizable phenomena, it is tempting to distinguish between metaphysically and ontologically irreducible cases. But for practical purposes this may not be the best way to proceed. As we saw in an earlier chapter, there is always the possibility that what we take to be a failure in our own knowledge may in fact be the result of an ontologically real indeterminacy (as in the case of the two-hole experiment). Moreover, it is difficult to say whether cases where an eyewitness codicil applies are really reducible in a pure sense (even though we can identify a unique basis cause), and if they are not, whether they are irreducible in an ontological or a metaphysical way. (Recall again Darwin's letter to Gray: the Englishman *happened* to know the raw numbers which his American colleague had used to reach a value of two percent, but absent that eyewitness codicil, the percentage value would have remained irreducible to a unique ratio of raw numbers.)

## (2) A brief taxonomy of reduction scenarios

To circumvent these practical difficulties, we can distinguish between "opaque" and "transparent" relationships. The meanings of these terms are probably obvious. If we have a practical means of reducing a situation to a unique causal basis, then the relationship is transparent, and if not, then it is opaque. Now it appears that many recursive contexts are transparent in an immediate sense, but opaque with respect to their initial conditions. (Recall what was said above about layered multiple realizability: We may be able to reason through some of the layers but not all.) Although the basic concepts addressed so far are not complex, it may help to distinguish explicitly among four kinds of phenomena. To this point we have touched on circumstances which are:

(1) actually reducible;
(2) epistemologically reducible;
(3) actually irreducible;
(4) epistemologically irreducible.

This distinction seems to be general, so certainly we can apply it to the class of all multiply-realizable phenomena. Furthermore, it is immediately clear that these categories are not mutually exclusive. For instance, Rosenberg's argument (1985, 1994) for the independence of biological science is based on the claim that propositions in biology belong to classes (1) and (4); that is, the phenomena studied by biologists conform to the rules of physics and chemistry, but we cannot perform an exhaustive reduction of all claims in biology to basic contentions of physics because of our limited conceptual apparatus.

Moreover, the third category *seems* to entail the fourth: if it is metaphysically impossible to attribute a given phenomenon to a specific cause, then it must also be true that we cannot find such a causal link. Someone might claim to have uncovered the specific agency involved in a given metaphysically intractable circumstance, but such an assertion would be erroneous and would therefore not count as an instance of real epistemological reduction. The difficulty Darwin points out to Gray, for instance, *cannot* be solved *in general*. It is easily demonstrable that a given percentage value can be converted into a ratio of the form x/100, and that the ratio can then be expanded into an infinite set of equivalent fractions of the form (y*x)/(y*100). Given the set of all such ratios, there seems to be no privileged linkage. If through some accident Gray's raw observations had been lost, there would be no way to reproduce them given only the clue that native species belonging to the Umbelliferae constituted two percent of native plants generally in a specific area. Anyone who claims to have overcome the *general* metaphysical difficulty here must be mistaken, so in fact no *general* epistemological reduction can take place. But we must remember the special kind of knowledge which we called an "eyewitness codicil" above; in essence, that is what a peek at Gray's notebooks would amount to. In fact, some phenomena which are normally irreducible may actually be attributable to a single, definite cause if an observer happens to be in the right place at the right time.

### (3) Another way of emphasizing the irreducibility of recursive contexts

That reflection and a moment's scrutiny of the four categories above tend naturally to raise again two questions which we have already encountered: (1) Are there any metaphysically irreducible entities "in the real world," or do such things occur only as the creations of mathematicians? (2) Are there any metaphysically irreducible entities in any realm which are not reducible in specific cases under an eyewitness codicil? Once again I want to avoid answering these questions, but it will be helpful to our understanding of reduction to probe a little around their edges.

It is easy to fabricate a "one-way" (irreducible) multiply realizable entity such as a percentage value, but it is not immediately clear from everyday experience that there also exist objects or circumstances in the world which are similarly one-way with respect to their causes. If there were such things in our world, it would be difficult even to describe our means of recognizing them, just as it was difficult to imagine what a murder mystery based on real metaphysical rather than apparent epistemological irreducibility might look like. Clearly we can conceive the possibility of *epistemological* irreducibility outside the realm of mathematics where chance is operative. When the establishment of causal links in the real world is especially difficult, whether for epistemological or ontological reasons, a classic way of describing the situation is to employ opposition or negation. Aristotle, for instance, deals with irreducibility when he speaks of circumstances which are *not* what we experience "always ... or for the most part" (*Physics* B.5, 196b10; Apostle (tr) 1980: 33), or in other words "uniformly or generally" (ibid., Wicksteed and Cornford (tr) 1980: 147)[16]. It appears that for Aristotle normality -- what occurs "always or for the most part" -- is metaphysically reducible to its physical causes. He does not possess a unique vocabulary to distill what does *not* normally occur, so he reverts to opposition or negation, that is, to saying what is *not* the case and how the happenings in question are contrary to the usual, tractable course of things. The class of circumstances thus described through negation is of course irreducible, since its cause is not a single, definite event nor a delimitable confluence of occurrences, but is rather luck or chance (τυχη or αυτοματον). This approach to irreducible multiple realizability is perhaps not surprising when we reflect that in the history of philosophy a common approach to transcendent agency -- which is what luck and chance represent in this case -- is a *via negativa*. That is what we find in Maimonides, for instance, as he reflects on the available means of talking about God as a wholly transcendent agency. Thomas Aquinas, too, employs a negative approach in addition to an analogical method.

Again, however, we must bear in mind the distinction between epistemological and ontological reducibility. Aristotle views luck and chance as agents of happenings which cannot be attributed to a definite cause. But the Philosopher's analysis arguably leaves open the possibility that the irreducibility involved in the agency of luck and chance stems from the observer's ignorance rather from a metaphysical indeterminacy inherent in the conditions themselves.

> "For example, a man engaged in collecting contributions would have gone to a certain place for the sake of getting money, *had he known*; but he went there not for the sake of this, and it is by accident that he got the money when he went there" (*Physics* B.5, 196b34; Apostle (tr) 1980: 34; emphasis added).

Of course this is not the only example of luck and chance which Aristotle offers, but arguably none of his examples indisputably deals with workings of chance and luck which are ontological rather than epistemological.

By contrast, the *via negativa* employed by Maimonides, Thomas, and others is committed to the unknowability of God as the consequence of an unbridgeable metaphysical separation. However, there is no corresponding acknowledgment of multiple realizability within the sublunary realm -- where questions of fitness presumably must be posed -- and so it is not clear that we can profitably employ a *via negativa* or *analogia* in analyzing fitness. In contexts where these methods apply, a given phenomenon may be the work of an unknowable agency, in which case all but stochastic phenomenologies are beyond our abilities. It follows that we cannot employ human language (including symbolic systems such as mathematics) to discuss the act of agency involved. Despite God's simplicity, however, we may conceive of a kind of epistemological multiple realizability involved in how God acts: maybe God did what He did for one reason, maybe for another. For instance, when Christ poses the famous question, *lema shabachtani*? ("Why have you forgotten me?"; Matt. 27: 46), the irreducibility of God's intention from a multiplicity of possibilities arguably reflects human ignorance of the transcendent rather than a metaphysical capriciousness. But once again, there would have to be a crossing of the sub-/ superlunary border before we could know for sure.

So the question remains, Does metaphysical irreducibility exist outside the realm of mathematics? As was already mentioned, one approach to the puzzle would be to argue on the basis of something like the famous two-hole experiment (Feynman 1965: 127 - 148; Gribbin: 1984: 164 - 176) that there are cases in which physical objects behave with a kind of randomness which forecloses any possibility of attributing definite causes to their motions. In other words, some "real-world" phenomena -- things which exist independently of our creative ability -- seem to be multiply realizable and also irreducible to a single, specific cause in an ontological sense. Interestingly, the analysis of quantum phenomena seems to rely on a combination of negation and analogy which we have already seen in what might be called the "pre-scientific" philosophy of thinkers such as Aristotle, Maimonides, and Thomas (meaning no disrespect to Aristotle, the student of biology, astronomy, and physics, nor to the physician Maimonides, whose scientific interests were nearly as broad as Aristotle's). Feynman sounds Aquinian when he introduces the two-hole experiment in this way:

> So then, let me describe to you the behaviour of electrons or of photons in their typical quantum mechanical way. I am going to do this by a mixture of analogy and *contrast*. If I made it pure analogy we would fail; it must be by analogy and *contrast* with things which are familiar to you (ibid: 129 - 130; emphasis added).

We can read Feynman's word "contrast" as "negation": he means that he will describe quantum phenomena by analogy and by saying what they are *not*, how they do *not* behave. Had he suggested that his mostly lay audience could only understand a presentation couched in analogical language, we might infer merely that the mathematics and physics involved were too technical and therefore had to be translated into terms familiar to his audience. But Feynman's partial reliance on negation (contrast) indicates not the difficulty but rather the impossibility of positively translating the empirical findings of quantum physicists into parlance referring to objects of common experience and how they behave. It would be suspect from the standpoint of logic to claim that the use of negation as a tool of rhetoric necessarily implies irreducibility, but it is clear from Feynman's presentation that the use of negation is a symptom of a metaphysically real causal indeterminacy. He takes care to stress that some quantum phenomena are not unpredictable because we are ignorant of some particular bit of information in the way that Aristotle's imaginary lender was unaware that a debtor would be at the marketplace. The notion that our ignorance of how metaphysically random phenomena behave results from incomplete knowledge of initial conditions

> "...is called the hidden variable theory. That theory cannot be true; it is not due to lack of detailed knowledge that we cannot make a prediction" about how certain particles such as protons and electrons will behave (ibid: 146).

This chapter is not primarily about quantum physics, but Feynman's reflections are important for our present purpose in a way which can be summarized as a kind of moral: the notion that propositions in biology can be theoretically reduced to claims of physics does not entail that biological organisms behave predictably in the sense that their emergent properties can be reduced to specific physical circumstances. At some point there is a discontinuity between the two realms. In one, reduction of specific causes to specific effects may be possible in a metaphysical sense, so that the only irreducibility is epistemological (Rosenberg's position on the parochialism of biology); in the other, it is not clear that any knowledge of initial conditions and operative forces, no matter how complete, can allow us to say with certainty that among all possible causes, a specific one was in fact the agent of a given effect. To the extent that the movements of organisms may be theoretically describable in terms of movements of elementary particles, there would still be an unpredictability in changes among the macro-phenomena corresponding to the unpredictability of movement among the elementary particles. That unpredictability amounts to irreducibility of a kind.

But it may be argued that by this reasoning irreducibility depends on the metaphysics of particles, while on a macroscopic level no such metaphysical indeterminacy exists. Rather (the argument would continue), failures of reduction and prediction on the level of organisms and even their DNA are due only to a lack of knowledge. If this were granted, then the argument would conclude that the fitness of organisms would always be reducible to concrete circumstances provided our knowledge were sufficiently complete. But this conclusion would be false, even granted the premises. Once again, what counts is the nature of the function which relates quiddities in such a way that some are supervenient on others. Let us tentatively take the following proposition as a principle:

In the case of some *recursive* functions, a unique reduction cannot take place simply *because we cannot always be certain where the reduction should end.*

Expressed a bit differently, for some recursive functions any finite reduction will appear to be what we have called a "broken elevator" reduction above, since (again depending on how the function is expressed) there need not be an exit condition going backward and every stopping point is premature from some perspective.

Consider the recursive procedure which expresses Wirth's solution to the eight queens problem, for instance. There is a clear exit condition in the "forward direction": as soon as we run out of squares or place the eighth queen, the procedure ends. But if we are given a board occupied by X queens and told that it was set-up using Wirth's solution algorithm, we cannot say with certainty how the present scenario's development began. There are 64 possibilities (absent an eyewitness codicil), none of which is privileged unless we make assumptions (e.g., how a human programmer would probably choose to begin a trial-and-error method of placing the queens) not explicitly contained in the solution algorithm itself. In that sense, the question of whether irreducibility is metaphysical or epistemological loses urgency. What counts is that when we attempt to reason "backward" from a present circumstance with the aid of the algorithm which created that state of affairs, we may find that there is no unique initial condition associated with the algorithm. That is true even of some patently practical algorithms, where one criterion of practicality is the presence of a fool-proof exit condition when executing the algorithm "forward."

It seems to me to be clear without the need of a lengthy argument that any extant way of calculating fitness values *can* be formulated as a recursive algorithm (function) of the following form

$$\text{Fitness}_t = f_i(\text{Fitness}_{t-1})$$

where the function $f_i$ takes account of particular variables found in the present taxon or individual as well as the selective environment which were not present at some previous stage. That odd phrase, "previous stage," takes its specific meaning from the purpose at hand. For instance, we might be interested in a "snapshot" of a species' fitness in a selective environment at a given moment in the past, in which case "previous stage" refers to a range of calendar dates. Alternatively, we might find it more interesting for a certain purpose to derive present fitness as a function of the previous generation's fitness, regardless of when that generation flourished. Or we might wish to skip several generations or even millions of them and inquire what happened to organisms during a "slice" of their cladistic progress.

Once again clarity may be served by reflecting on what the foregoing discussion does *not* conclude. Kantian terms are convenient for the purpose. There has been no general siege of the notion that all synthetic *a priori* judgments are necessary nor a specific attack on the claim that every change has a cause. This is not to say that such an assault could not be made, only that it has not been our aim above. Instead, we have investigated whether the cause of a specific, given event is always *identifiable* as a single, definite phenomenon among the range of possible causes which we attempt to distill through whatever algorithm we think led to the status quo (cf. Körner 1955: 24 - 32; Friedman 1992: 246, 250). The answer appears to be that in the case of some recursive functions, a unique reduction simply is not possible, and not just due to a hidden variable. The problem is that a recursive development of some multiply realizable scenarios precludes finding a single, unique set of initial conditions.

(An interesting aspect of some rejections of a thoroughgoing creationism -- the view that everything now present was, at least in its form, created by a single agent rather than having evolved from a plethora of circumstances -- is that at least some if not all present phenomena are either too multifarious or too imperfect to be the result of a single design or designer. Von Ditfurth, for instance, cites Kant's argument from the *Allgemeine Naturgeschichte* (apparently referring to 1755: II.8, 331 - 347) that whatever we may believe about the ultimate origins of the universe's "stuff," tiny divergences from regularity argue for the *evolution* of things like planetary orbits rather than their divine *creation* (von Ditfurth 1982: 38 - 39). If we conceive evolution as an essentially recursive process, in which a set of circumstances acts upon and subtly modifies the world and its inhabitants, which in turn influence the selective environment acting in the next "iteration" of the process, then we can read arguments such as Kant's and von Ditfurth's as validating our conclusion that a recursive function need not indicate a single set of initial conditions.)

## (4) Reducing to things, reducing to propositions

At this point it would be well to draw yet a further distinction, but this time one parallel to those already drawn. When we speak of reduction, it appears that we sometimes mean the reduction of one entity to another, but at other times we may mean the reduction of one proposition to another (or to a set of propositions). Let us call these two types "object reductions" and "claim reductions," respectively. The sense of the nomenclature becomes clearer when we attempt to assign examples to the two categories. Reducing a murder victim to a murderer does not mean that there is some alchemy by which one body is transmogrified into another. Rather, a multiplicity of possible causes expressible as propositions of the form, "The judge was the murderer," "Davis was the murderer," etc., can be reduced to a single proposition which, among all the conceivable assertions "X was the murderer," is unique in that it alone is true. Murder mysteries thus involve claim reductions rather than object reductions. What, then, is an example of an object reduction? If we stipulate that reduction need not be unique, then we can say that the percentage value which engaged Darwin's attention can be manipulated with an object reduction -- we can show it to be equivalent to any number of ratios. But the qualification of uniqueness has been the implicit *sine qua non* of reduction as we have understood the term up to this point. In other words, we tend to think that attempts to reduce multiply realizable phenomena are successful only if the reducing algorithm yields a single result, whereas a weak reduction (which merely narrows the range of causal possibilities associated with an existing state of affairs) is not normally considered to be a reduction at all. Let us stipulate that henceforth, "reduction" and its cognate terms mean unique reduction unless otherwise indicated. Given that convention, it is not clear that we can find any examples of pure object reduction. That is partly because we can always embed an object in a proposition (e.g., rather than saying that we reduce two percent to 2/100, we can say the claim "The value of that function is two percent" reduces to the proposition "The value of that function is 14 in 700"), whereas it is arguably impossible to conceptually embed a proposition in an object. Moreover, the objects we have so far considered -- including ratios in the form of percentage values and first derivatives, and photons and electrons as they behaved in the two-hole experiment described by Feynman -- are all irreducible in one sense or another. In the former cases, reduction is possible only if an eyewitness codicil applies; by contrast, there seems to be a metaphysical irreducibility in some quantum phenomena.

It is interesting to note a related phenomenon discussed by Quine in his indeterminacy of translation thesis (ITT). His contention can be summarized in the following passage:

> Yet one has only to reflect on the nature of possible data and methods to appreciate the indeterminacy. Sentences translatable outright, translatable by independent evidence of stimulatory occasions, are sparse and must woefully under-determine the analytic hypotheses on which the translation of all further sentences depends. To project such hypotheses beyond the independently translatable sentences at all is in effect to impute our sense of linguistic analogy unverifiably to the native mind. Nor would the dictates even of our own sense of analogy tend to any intrinsic uniqueness; using what first comes to mind engenders an air of determinacy though freedom reign. There can be no doubt that rival systems of analytic hypotheses can fit the totality of speech behavior to perfection, and can fit the totality of dispositions to speech behavior as well, and still specify mutually incompatible translations of countless sentences insusceptible of independent control. (Quine 1960: 72)

It is worth noting as well that variations of Quine's insight (or Duhem's, some might well claim) are still being used to argue for various ontologies.[17] Quine's observations serve to strengthen our contention that scenarios derived from recursive algorithms can be irreducible. In general terms the ITT is simply a special case of irreducible multiple realizability. But we may well wonder about the phrase "totality of speech behavior" and "totality of dispositions to speech behavior." What do such *totalities* consist of? If we really knew *everything* involved in a speech act, no matter how indirect that involvement might be, could we not then arrive at a unique translation? Similarly, if we knew enough about a state of affairs and the recursive algorithm associated with it, could we not achieve a total reduction? I believe the answer is no -- that Quine had it right -- but in the next section I will try to argue the opposite side. At the end of the section, we will see why the argument does not work.

## (5) Cumulative reducibility.

Here we confront the paradoxical possibility that some concatenations of metaphysically irreducible entities can become reducible, both in fact as well as thought. Consider a first derivative in two-space -- a vanishingly small change measured against an ordinate compared with a vanishingly small variation based on an abscissa. Described merely as a numerical value, a specific derivative may describe a point on any number of actual or possible curves, anything from the changing surface of a wave on the Neckar River in Heidelberg to a portion of the art-deco embellishments on the Chrysler Building in New York City. In this sense the first derivative is pretty much like the percentage values which Darwin, as a naturalist interested in causal reductions, found insufficient for some of his purposes: again, the derivative is a ratio (only this time of limits), a mathematical formalism; it is arguably not a "real" part of the natural world at all. Moreover, it is multiply realizable and also irreducible, both metaphysically and epistemologically (unless an eyewitness codicil applies), to any single curve instantiated by a real phenomenon.

There is, however, something bothersome in the assertion of irreducibility. One way to isolate the annoying element is to play a game involving excluded middles. The Chrysler building has a rather unmistakable profile, and presumably it possesses the only such contour in the world. Assume (for simplicity's sake) that we consider the profile alone, as a two-dimensional image, and then take a first derivative at some point along the surface of that silhouette. We present the resulting value to an expert in the history of urban architecture after carefully explaining the units of measurement involved, and then we ask the specialist what the number represents. It would be miraculous indeed if the authority responded: "Hmmm...rings a bell....Hey, that number, understood as a first derivative, could correspond to a curve constituting part of the profile of the Chrysler Building viewed in cross-section!" But rather than hoping for miracles, we make the game easier. We tell the expert that the number provided is a first derivative of the line constituting a famous building's silhouette. Of course even a learned student of architecture could not discern a pattern from a single derivative, so we begin to increase the number of derivatives in the expert's data base and offer them in order. We even plot the evolving contour after each addition. For some sufficiently small number of "information sets," {first derivative, order in the sequence of data points, cumulative visual representation}, each set as well as the sum of sets remains irreducible. But at some point in the process it would also be miraculous if the historian-architect did *not* recognize the Chrysler Building.

Under these circumstances, it appears that one information set is irreducible in the sense that it cannot be linked to a specific architectural profile. Two information sets are similarly irreducible, as are three, and perhaps even three thousand, depending upon their distribution. But at some point, N information sets become reducible when evaluated by a given algorithm (which -- with no offense intended to historians of architecture in general -- is what the analyst in our example represents). We would arrive at the same conclusion if we talked about portions of curves: as we increase the number of samples, patterns become suddenly recognizable and some of them become uniquely tied to a set of samples. The reducibility may not be perfect, since it is probable that there are many pictures of the Chrysler Building which the collection of derivatives might represent as accurately as they do the building itself. If we changed our example so that it dealt with three rather than two dimensions, we would then have to contend with the correspondence of three-dimensional models of the building with the derivative-based picture. But probably none would dispute that the accumulation of individually irreducible values can result in something which is at least weakly reducible. Let us call such a phenomenon "cumulative reducibility." What this amounts to is a limit of the range of possible correspondences.

But even if cumulative reducibility is a fact, it does not occur necessarily and so does not contradict our earlier conclusions about the irreducibility of recursive algorithms. For instance, Darwin would not have been satisfied with a huge number of congruent fractions. In that case, accumulation does nothing to solve the problem of irreducibility. The view of fitness defended in this dissertation -- that of a recursively defined property of an organism (or indeed of any taxon) -- entails that only intra-definitional reductions are possible. This means that we can reduce along the "chain" of recursive accumulation from, say, $p_k$ in the reading-assignments example (chapter 3, 3.(2)) down through $p_{k-1}$ to $p_{k-2}$ and so on, but eventually what we reach is $p_1$. That point remains a part of the pedagogical algorithm, not outside of it. We have no way of knowing whether that point was the initial condition of the current algorithm or not, unless further information (i.e., information not provided by the algorithm itself) is given, hence we cannot know that we have reduced to an initial condition.

Suppose that as observers of what transpired in the classroom we had no knowledge of the recursive algorithm being employed. We simply treat the phenomenon of repetitive teaching in the classroom as a behavioral phenotype of the organisms involved (or even as an extended phenotype, to use Dawkins' terminology from 1985). After a sufficient period of observation a bright ethologist somewhere is sure to discern the recursive pattern. Moreover, if this recursive pedagogical method were sufficiently widespread and of long standing, we would assume that it probably has an adaptive advantage. Under such circumstances, to what physical circumstances would we or could we reduce the recursive algorithm? The answer depends on what we mean by "*the* recursive algorithm." If we mean the phenomenon in which a given, well-defined group of students read the first page of a specific work of Kant on a certain day of a particular year, then read that page and the next on the following day, etc., the answer is that the description of the algorithm is itself a reduction of sorts. But if we represent the key term of the algorithm more abstractly, e.g., as something like the verbal description given above or as an expression such as $\sum_{i=1}^{n} p_i$, then the appearance of reduction can be deceptive. (It is important to realize that this term is not the algorithm itself.)

An algorithm *qua* method presupposes questions such as (in this example):

(1) What reading shall the students be assigned on a given day?

(2) How many pages (*not* including repeats) have been read on a given day D (where D is understood to be a natural number corresponding to the number of days of

instruction following the first one, so that the D corresponding to the first Wednesday would be 3)?

(3) How many pages (including repeats) have been read on a given day D?

It is immediately obvious that question 1 is forward-looking while questions 2 and 3 are backward-looking. Moreover, question 2 could be solved recursively, but that would be needlessly complicated; the answer is obviously D, that is, the same number of pages have been read as there have been days of instruction. Question 3, on the other hand, almost begs for a recursive solution. It seems natural to solve it using something like this algorithm:

```
DAY = <current day>                    (initialize)
PROCEDURE  Cnt-pp(n:  INTEGER):  INTEGER
    IF DAY  = 0 THEN RETURN 0
    ELSE RETURN Cnt-pp(DAY - 1)
        DAY = DAY - 1
    END
END Cnt-pp
```

There are many things we can say about the relationship between this algorithm and the concrete circumstance occurring in the classroom of an academic culture which values repetition. We can say a certain algorithm *describes* the actual situation, or *applies* or *corresponds* to it, but we cannot say that it *reduces* to the situation. Viewed from the opposite perspective, the concrete circumstances can be used to exemplify the algorithm but not to define it. To see this, it helps to realize that an algorithm is a means of predicting (in forward-looking circumstances) or reconstructing (in backward-looking situations) *part* of an occurrence. An algorithm does not "calculate" -- either in the sense of predicting or reconstructing -- the happening as whole. We might call that a quantitative argument. If the domain of A does not equal the domain of B, then A and B cannot be the same.

But we can arrive in a much more general way at the conclusion that algorithms are irreducible to the concrete circumstances to which they apply or correspond. To do so we first need to accept that circumstances and algorithms used to predict or reconstruct specific aspects of them are not in the same genus. When things are of the same basic type -- when they are defined in like terms (e.g., qualifiers belonging in the same Aristotelian category) -- then we expect that it is possible to translate one term into another. That is because we are dealing with a continuum of sorts. Regardless of how widely the terms are separated from one another, even if they are polar, we can always "get there from here." Kipling's famous poem (1890), "The Ballad of East and West," asserts: "Oh, East is East, and West is West, and never the twain shall meet...." Kipling means, of course, that there is always a gulf between the Occident (whose archetype for him is Great Britain) and the Orient (represented here by the area straddling what is today the border between Pakistan and Afghanistan). The line makes for high drama and interesting social commentary, but it's nonsense unless taken metaphorically. Of course the "twain" meet, else a great deal more shipping would be lost. The line comedian Bob Hope sung in the film *Paleface* is logically more defensible even when perceived in the most simple-minded way: "East is East, and West is West, and the wrong one I have chose...." The line makes sense even when taken literally. But algorithms and the real-world circumstances to which they correspond are not on the same continuum. They, like Kipling's metaphorical East and West, do not meet. That in turn entails that the one is not reducible to the other in any literal way.

Let us agree for the sake of simplicity that a normative statement of the form "Students will read X number of pages per day, including Y repeats" is actually a species of predictive algorithm. We can also predict the number of pages a student will have read on a certain day, and we can reduce the one algorithm to the other. But we cannot say that either prediction is in any way the same as the act of reading itself. In fact, the gap separating a state of affairs from an algorithm used to predict or reconstruct it mirrors a classic sort of dualism -- the kind, for instance, that separates what Dennett calls Cartesian "mind-stuff" from patently physical processes taking place in the brain and elsewhere in the nervous system. Dennett suggests the trouble with this kind of dualism is that it does not allow us to find continuous, mechanistic explanations for phenomena such as consciousness. If we separate mind and brain, calling the one non-physical and the other physical, we have no way of explaining the interaction of the two realms (1991: esp. 33 - 39). The point of this section is not to take sides on the debate which exercises Dennett, but we can appreciate his point in the abstract and apply it to our present concern. If we agree that algorithms and the portions of "real-world" circumstances which they describe are of essentially different kinds -- if placing them on the same continuum would be what Ryle calls a "category mistake" -- then there is no possibility of reducing algorithm to circumstance. So it is with fitness as well. If fitness is conceived as a formal recursive algorithm, it is not reducible to specific circumstances (through an object reduction) nor even to a single, identifiable "initial" state described as a proposition.

## (6) Taking stock

We know that some phenomena are reducible while others are not. Here, reducibility is used to specify the existence of a causal link which unites a unique cause with a given effect. If the link is necessary in a way independent of all observers, then we say the phenomenon which plays the role of effect is ontologically (or metaphysically) reducible to its cause. A metaphysically reducible phenomenon may not be epistemologically reducible for a given observer, which is to say, the analytic powers of the observer may not suffice to unpack the causal connection involved. If the link between cause and effect is sufficient (i.e., if the presence of the cause suffices for the presence of the effect), then we are confronted with a problem of nomenclature. If we happen to have knowledge of the link in question through some means -- if Gray happened to have included in a letter to Darwin the raw numbers corresponding to a certain percentage value of introduced plants as a proportion of indigenous plants or if I chanced to see that the sensation I felt was due to a neighborhood cat playing with my shoelace and not to some other cause -- then it seems correct to say that the value or sensation in question is epistemologically reducible in one sense because the connection can be accurately specified under an eyewitness codicil. But if it the connection is epistemologically specifiable, then it must also be metaphysically describable. Let us say that in such a case the status of the causal link can be called "coincidentally reducible" but "generally irreducible." But consider what would happen if the phenomena in question were formally defined, with no extraneous information provided. The relationship of a percentage value to the set of ratios is given by $p = (x*p)/x*100$, where x is any rational number, and any p is irreducible. In the case of a recursive function, the prospects of reduction are even more hopeless because recursion implies multiple iteration,but there is no way of telling how many iterations separate the present circumstance from the initial conditions. Moveover, the conditions and the algorithm belong to separate genera.

That word -- genera -- needs to be stressed. To this point we have treated membership in separate categories of existence as being one of the obstacles to reduction. It should be noted that reduction can fail on this score in an absolute way. It is not that the science of one epoch attributes a phenomenon to one causal basis and then later corrects itself by changing the realizing basis.

Consider a quality such as foot speed and its relation to physical properties. A few decades ago it might have been considered an engineering problem: mechanically speaking, some macroscopic proportions -- say of leg length compared to overall height -- allow an organism such as a human being to run faster while other proportions lead to a slower gait. But times change, and the basis for making such distinctions is now explored on what could be called a more basic level. In 1992 the most popular running magazine in the United States published a cover story entitled "White Men Can't Run." The conclusion was that West African blacks make outstanding sprinters based on their high percentage of "fast-twitch" muscle fiber. East African blacks, by contrast, possess relatively ideal attributes for long-distance running. Stuck in the middle are "hybrids" possessing a mix of both attributes. A sample of the *Runner's World* article makes the method of reasoning clear:

> On average West African blacks have 67.5 percent fast-twitch muscle fiber and French Canadian whites 59 percent. Given a normal distribution curve, there should be more black individuals than whites at the far end of the curve where Olympic sprinters would be found. (Burfoot 1992: 93)
>
> When running a marathon on a laboratory treadmill, South African blacks were able to use a higher percentage [90%] of their max VO2 than average [75%], good [80%] or elite [85%] white marathoners. (ibid.: 94)

The explanatory goal of the article can be described as reductive: it sought to explain why a roll call of recent olympic running champions reveals what seems to be a disproportionate number of black athletes (disproportionate, that is, given the racial composition of the competitors in general). A few paragraphs were devoted to sketching past explanations in which a much different reduction had been made, one emphasizing environmental factors to the total exclusion of physiological differences.

To repeat, that kind of debate is *not* among the obstacles to reduction which we have considered. We have not been interested in the fact that radically different explanations of a particular adaptive behavior may be possible. Rather, in saying that a "genus gap" makes reduction impossible, we should understand a much more foundational difference. It is incorrect to say that fitness cannot be reduced to concrete circumstances simply because we see that causal explanations for factors relevant to fitness (e.g., running speed and endurance) have been revised. Even if we were positive that we had correctly defined the concrete causal basis of speed and endurance, we still would not necessarily have reduced fitness to those qualities, even in a "microcontext." That is because we can choose to define fitness in a formal way, thereby putting it in a wholly different genus from that of concrete qualities such as speed and endurance.

This chapter was billed as an attempt to clarify certain issues of reduction and supervenience which will be important as we develop a concept of fitness below, and so it is worth our while to try out every exploratory tactic as we try to analyze reducibility and supervenience. Taking a clue from Plato, Aristotle and a long tradition encompassing the so-called plain language philosophers, it may pay us to ask why certain notions like derivatives and percentage values (which Darwin insisted should be reported along with the raw numbers of which they are a quotient with denominator normalized to 100) can be profitably employed in disembodied form, without our consistently and precisely specifying *what* is being compared to *what*, while the units of other ratios are routinely specified or at least universally understood in their contexts of usage. The answer may have to do with a movement traceable at least back to Comte:

> "The triumph of the positive spirit consists in the reduction of quality to quantity in all realms of existence -- in the realm of society and man as well as in the realm of nature -- and *the further reduction of quantity to ever larger and more abstract formulations of the relations that obtain between abstract quantities* (Lenzer 1975: xxi; emphasis added).

It seems clear that fitness is something like a first derivative in being emergent, primitive, and *operationally* irreducible in specific cases. We can speak of fitness values -- quantities which make sense only in comparative contexts but which have no units appended to them. In other words, fitness values (like derivatives) are reckoned as ratios but not expressed that way. We speak of quantifiable concepts such as velocity in terms of miles per hour, meters per second, or any number of possible units. Occasionally we drop the units: "The cop gave him a ticket 'cause he was doing 70 in a 20 zone." But even if that were the norm, the listener could presumably say exactly what units are implied.

Perhaps it is the case that when the units of a quantifiable attribute are routinely dropped within a linguistic context -- e.g. when a racecar driver recounts to his teammates that he spun-out while doing "145" (instead of "145 miles per hour") or when a biologist names the fitness values of experimental subjects as similarly disembodied numbers -- it is precisely because the conversation takes place at a level of abstraction where the interest is in *comparison* rather than *reduction.* Once the driver and his listeners are thinking in terms of the same units, those units arguably have no bearing on the issue at hand. What counts is that on the previous day the disembodied number "135" was associated with the point of the curve where today's mishap at "145" took place. Those who believe in the reducibility of fitness will gleefully point out that in the racecar example a reduction may be possible even if it is not of immediate interest. In other words, the interlocutors *could* rephrase their every statement to include a unit such as miles per hour and thus reduce a proposition about the mishap to a proposition about a specific distance being covered per unit time at the moment of the accident. The analogy loses its force when we recall that we are considering fitness to be defined not just by *any* formal function, but rather by a *recursive* one. If velocity were defined not in terms of distance per unit time, but rather as a function of itself, then the discussion at the garage would include statements such as "Do no more than 93 percent of today's speed and you'll be alright on that curve." In other words, the possibility (not to mention the need) of reducing to concrete circumstances would evaporate. The objection that speed *must* be defined as distance covered per unit time will not hold up in such a case. A racing team in the Centauri 7 solar system might never have perceived their environment in that fashion or perhaps cannot have done so. For members of that linguistic community, speed might have to do with perceivable changes in the fluid of their inner ears. In any case, a wholly formal, self-referential definition need not be reducible to concrete circumstances.

To this point the term "reduction" has been used with so many possible meanings that we may suspect it actually refers to quite different phenomena. For instance, Darwin's remark to Gray is motivated by the fact that *numerical* reduction from a percentage value to a single definite ratio is impossible. Douglas's theory that insect wings evolved from structures whose adaptive value was thermoregulatory deals with the selection of one among several theoretical alternatives and is therefore a kind of *explanatory* reduction. The discussion of first derivatives as irreducible and the accompanying illustration using the Chrysler Building seem to typify the problem of numerical reduction, only with one qualification -- in the problem of numerical reduction confronting Darwin, there is no already-existent standard of truth which the analyst can use to solve the problem. A given percentage corresponds to uncountably many ratios, period, and he would have had to travel to North America to find out the truth of the matter in terms of raw numbers. This intractable multiple realizability applies even more strongly in the case of first derivatives in isolation. By introducing an already-known standard of comparison such as the well-known profile of a building, on the other hand, a problem of *comparative* reduction becomes theoretically tractable.

It need not concern us here whether this brief taxonomy of reductive problems is complete. For present purposes it is sufficient to realize that the three sorts of reduction discussed so far -- numerical, explanatory, and comparative -- appear to be common challenges confronting the evolutionary biologist. In some contexts we will indeed need to treat each kind separately, but we are justified in placing all three types under the heading of reduction in general. Looked at from the point of view of set theory, each of the problems has a form which we might call "one-to-many" (or just "one-many") and in each case the reductive agenda is to reduce the situation to a one-one correspondence if that is possible.

## 3. Supervenience

Let us consider a simple use of the term "supervenience." In the course of a book which does not presuppose any formal training in philosophy, Russell discusses the doctrine that mind and body are wholly separate from one another, a position based on but more thoroughgoing than Descartes' theory of separation. Describing a version of this doctrine associated with Geulincx and Malebranche, Russell writes:

> "To account for the relation we must suppose that the world is so preordained that whenever a certain bodily movement takes place, what passes for the appropriate mental concomitant does in fact supervene at the right moment in the mental sphere, without there being a direct connection" (1959: 197).

Here to "supervene" means more or less merely to "occur correspondingly," to come into the picture at a time corresponding to some other noteworthy course of events. What counts in such a sense of the term is the *venire*; the *super* has little if any meaning because in fact the mental event is not metaphorically above, beneath or on the same level with the physical in Russell's example; the mental event has no connection with the physical circumstance upon which it "supervenes" other than a *temporal* one.

We have no absolute grounds for ruling out such a sense of the term, but let us stipulate that a "disconnected" sense of supervenient -- however valid it may have been for Russell's purposes -- is not what we will mean when we say that one level of occurrence supervenes on another. Rather, we will assume that there is always some real *causal* connection as implied, for instance, in Rosenberg's definition of supervenience:

> The relation of supervenience has been expounded in contemporary philosophy as a relation between properties and sets of properties that have the two following characteristics:
>
> 1. If a set of properties A supervenes on another more basic set of properties B, then no two distinct objects that share identical properties from the more basic set B can differ in the properties they share with set A.
>
> 2. If the set of properties A supervenes on the set B, then it may be that no property in set A can be defined [in terms of] or manageably connected to any set of properties in set B. (1985: 113)

I take it that the conditional identified as 1. amounts to saying that there is a *causal connection* between A's quiddity and the set of properties B, while 2. asserts the possibility that A is irreducible to B despite the causal connection. (Here the irreducibility is epistemological, as indicated by the participle "defined" and the adverb "manageably.")

It is worth inquiring first what Rosenberg (or the "contemporary philosophy" on which he bases his definition) means by "basic." For the moment let us assume that it has to do with concrete circumstances as opposed to general terms used to describe relationships of concrete things or (at an even less basic level) relationships among the relationships. A few examples are in order to demonstrate the forms such concrete-to-general relationships can take.

(1) Rent à la Ricardo

In his *The Principles of Political Economy and Taxation*, published in 1817, Ricardo[18] follows Malthus in treating *rent* as being essentially supervenient on other properties.

> Mr. Malthus, too, has satisfactorily explained the principles of rent, and showed that it rises or falls in proportion to the relative advantages, either of fertility or situation, of the different lands in cultivation, and has thereby thrown much light on many difficult points connected with the subject of rent, which were before either unknown or very imperfectly understood...." (Ricardo 1973: 272).

Let us suppose that the situation were really as simple as this passage indicates (which neither Ricardo nor Malthus actually held), so that rent is a function of just three variables -- status of cultivation, fertility, and situation. Even if the elements involved were so few, however, we still could not be certain of reducing an observed rent to a concrete combination of the three variables in the absence of a reducing *function*. Moreover, even if the function relating the variables to rent were given (as we knew, for instance, the solution algorithm for the eight queens problem), it still is not certain that a reduction could take place. That is because we might not know where to stop as we attempt to reverse the algorithm. To make these observations concrete, assume we had the following algorithm for relating rent (R), to the three variables of cultivation status (C), fertility (F), and situation (S): R = C*F*S, where each of the variables is Boolean and set equal either to 1 or 2 (the land in question either is entirely under cultivation or not, meets a minimum standard of fertility or not, and has a good or bad location based on some standard). Then an R-value of 8 clearly means the land is prime, while an R-value of 1 means it could be in downtown Chernobyl. But apart from an eyewitness codicil, there is no knowing what to make of an R-value of 4 or 2, each of which corresponds to more than one combination of variables. At best R is weakly reducible in some cases, meaning that if we observe an R-value of, say, 4, we can exclude one of the possible combinations of values for the variables ({1,1,1), (2,2,2)}.

## (2) Pornography

Another very interesting example of supervenience just happens to be a currently hot topic in social ethics, at least in the U.S.A. The struggle to define pornography amounts to an exercise in justifying claims of reducibility and supervenience. That these exercises have often been more frustrating than fruitful is evident in a famous pronouncement on obscenity (a concept closely linked to pornography, at least in American jurisprudence) from the bench: "'I shall not today attempt further to define [obscenity]...; and perhaps I could never succeed in doing so. But I know it when I see it'"[19] (Strossen 1987: 53, quoting from *Jacobelli vs. Ohio*, 378 U.S. 184, 197 (1964) (Justice Stewart concurring); cf. Hawkins and Zimring 1988: 20). Had the good justice wished to couch his sentiment in the terms of our present discussion, he would have been justified in claiming simply that the concept of pornography is *supervenient* in a multiply realizable fashion on a set of representations (still photos, films, literature), but that he has found no means of *reducing* pornographic works in general to unique and necessary sets of generalizations about those concrete representations. Apparently there was justice in Stewart's majority opinion in this sense, that no matter what basis he had stipulated, a thousand voices would immediately have raised counter-examples and suggested additions. Call depictions of bodily contact among nude people pornographic, and at least one of Rodin's best-known works will make you a laughing stock. Suggest that such a locus forms an exhaustive basis for all pornography, and a pro-"family values" spokesman somewhere is sure to point out that an unseemly film or two involving people who are fully clothed (albeit perhaps in rubber) has been wrongly omitted from the class of pornographic works. Preferring discretion to valor, we won't attempt a definition, either. But that still leaves several general issues to be discussed.

It will surely be objected that if pornography is supervenient at all, it is not so in the same sense that fitness may be called supervenient upon physical properties of organisms. That observation leads us to an interesting reflection on the meaning of supervenience as the concept relates to fitness. An exposition of the differences between fitness versus pornography as supervenient might be based on the notion of absolute versus comparative supervenience. A judge could evaluate a given magazine as being pornographic or not without thereby implying in any sense that the pictures in question are more or less pornographic than any others in the world. A judge somewhere in the world might find such a spectral evaluation of pornography germane to the bench's pronouncements, but that need not be the case. Sometimes depictions are pronounced pornographic, period, and the gavel falls. By contrast, when we say that A is fit we cannot avoid meaning that A is fitter than some B. Fitness must always be comparatively rather than absolutely understood. Does this fact affect the kind of supervenience which fitness can possess? Certainly. In the case of a comparative quality such as fitness, the multiple realizability which characterizes the supervenient relationship must also take account of the multiple realizability of competitor organisms' fitness levels. In other words, comparative supervenience relationships tend to be more complex than absolute ones because more variables are involved in the former sort than in the latter.

Another criterion for distinguishing types of supervenient relationships relies on the difference between dynamic versus static quiddity. Some allegedly pornographic presentations change over time independently of an observer's activity. Films and song lyrics exemplify this sort of "dynamic" pornography: a film is rolled and a song is sung, both necessarily across a certain time period. A static moment of either object as process at best indicates the content of the whole without reproducing it. By contrast, a pornographic picture and words on a page "move" only according to what a reader or viewer does. Arguably, then, a pornographic character can supervene on a passive state of affairs. This sense of what we can call *static* supervenience corresponds to a static kind of fitness reminiscent of Mills and Beatty's propensity interpretation (1979). Such *static* fitness must be conceived as a function of a phenotype (or set of phenotypes) in a given time interval during which *no* changes are occurring in the selective environment and the organism in question is not exercising any countervailing phenotypes. During that interval, the relationship of phenotype to environment can be considered to be unchanging. But we think of fitness in general as deriving from action rather than passivity -- from a struggle to survive and reproduce. The same dynamic character seems to apply to the data base of "track records" from which we make generalizations about fitness (e.g., "Foot speed generally improves the fitness of land-dwelling predators, *ceteris paribus.*") and formulate specific theories and predictions about fitness ("Under certain conditions the presence of a sickle-cell gene enhances fitness by affording partial immunity to malaria and yellow fever.").

So it seems we are dealing with two kinds of supervenient fitness, distinguishable from one another based on how and for how long the observer chooses to abstract phenotypes and conceptually "freeze" the interplay of organism and environment. Returning to our example of pornography, a judge may find a certain topless-bottomless still photo to exceed the bounds of decency (and any who disagree are implicitly pronounced *sansculotte* at the same time). In this case the supervenient pornographic character of the photo is passive. But presumably the judge employs a background of *active* experiences in order to make the ruling. She recalls *dynamic* situations associated with the dorsolithotomic posture, she imagines *active* employments of some of the phenotypes displayed. Perhaps we are dealing with a kind of layered supervenience: at one end of the spectrum, "microfitness values" are associated with very specific, active relationships of organism and environment over time; at the other end is a list of static phenotype-environment-fitness triplets inferred from the active background data. To test this model and to set the distinction between passive and active supervenience into sharper relief, we should try to find a patently active supervenience relationship outside the realm of evolutionary biology. If we conceive of human warfare as intentional strife with the immediate goal of surviving and then gaining dominance over one's real and potential opponents, we are fairly near a model of the interaction which produces fitness differences. The struggle to survive and reproduce in nature may seldom be intentional (as most would understand the term), but the similarities between the two models -- of human warfare and of the struggle to survive in nature -- are close enough so that the active components of each resemble one another.

### (3) Clausewitz's "friction" and Walzer's "supreme emergency"

Karl von Clausewitz (1780 - 1831), a Prussian general and renowned strategist, discussed a phenomenon which he claimed emerged unintended from the activities of battle. He called it "friction," and we might understand it as the unwanted and unintended confusion which results from a surfeit of activity combined with a dearth of data and time for reflection. Clausewitz's perspective was that of a commander expecting the unexpected, knowing that battle plans presuppose adherence to plan, which in turn assumes the timely receipt and assimilation of information.

But battlefield conditions generate their own dynamic. One need only open a book of military history, almost at random, to find copious restatements and concrete instances of the Clausewitzian concept of friction. Speaking of the intimidated Allied troops trying to create beachheads on D-Day, Ambrose writes:

> [T]he GIs who landed at the wrong place and whose officers were wounded or killed before they made the seawall did not know what to do next. Not even heavy gunfire puts such a strain on a soldier's morale as not knowing what to do and having no one around to tell him. (1994: 368)
>
> In general, the men cringing at the seawall were as confused as they were exhausted and shell-shocked. In the wrong sectors with none of their leaders present, they just did not know what to do. (ibid.: 428)

Naturally the lack of data is two-sided here: the soldiers don't know what to do (and therefore almost certainly are not doing what they were supposed to do), and their commanders, removed from the scene, do not know what is happening. In particular, the commanders do not know that their troops are confused, meaning the wrong decisions will be taken and the troops will become even more disorganized and ineffective, which will make the commanders' plans even more inaccurate as mirrors of the battlefield situation, and so on. Moreover, from the commander's perspective, the unexpected can occur for many reasons, and not just because soldiers are terrified and confused.

> By daylight, June 7, there had been no counterattack. What Washington did see at first light was astonishing enough: two GIs leading fifty German POWs into the American line. The Americans turned out to be privates who had been mislanded and captured by the Germans. Both American privates were of Polish extraction; the "German" soldiers were Polish conscripts; when darkness fell the GIs persuaded their captors to hide out in the bushes and surrender at first light. (ibid.: 447)

Students of military history and ethics encounter odd justifications for wartime activities based on supervenient criteria such as "supreme emergency," devised by a contemporary American ethicist to assert that the intentional bombing of German cities with the aim of killing non-combatants and destroying infrastructure not directly related to the Nazi war effort was justified "during the terrible two years that followed the defeat of France, from the summer of 1940 to the summer of 1942, when Hitler's armies were everywhere triumphant," whereas (so goes the rest of the argument) later attacks on non-military targets in Germany and Japan were unjustified (Walzer 1992: 255). The criterion of supreme emergency is made supervenient on two properties which themselves are supervenient: the *nature of the threat* facing one set of combatants (the "danger must be of an unusual and horrifying kind" and threaten "the survival and freedom of political communities" rather than just a number of individuals conceived as that and nothing more -- ibid.: 253 - 4) and the danger must be *imminent*. What we see then is an argument based on a layered supervenience: "supreme emergency" supervenes on the nature and imminence of a threat, elements which supervene on the physical circumstances of an ongoing conflict. The weaknesses and tensions in Walzer's argument aside, its form seems relevant to the way we reckon fitness. We consider that fitness supervenes on phenotypes such as behaviors which themselves are supervenient on more "basic" qualities (to use Rosenberg's term). One of the primary weaknesses in Walzer's argument is itself an argument against the reducibility of supervenient fitness: the "foundation" of concrete circumstances is so far removed from the supervenient concept of supreme emergency that the ultimate determination seems arbitrary. (Who's to say what constitutes "imminence"? What about a "danger of an unusual and horrifying kind"? In the absence of well-defined standards, any answers to the questions will boil down to Stewart's pronouncement on obscenity: "I know it when I see it.")

It is important to reflect that the cumulative, layered character, emerging from non-deterministic levels of dependence, lends itself to expression in a recursive algorithm. Take the case of pornography. The issue is not one of relating a physical state in one area of a body to a physical state in another area, as when stimulation of the vaginocervical area in rats demonstrably increases the rate of glucose uptake in various areas of the brain (Allen *et al*, 1981), although that situation may also represent a supervenience relation. Rather, we cannot reason all the way back from an object which is intuitively pornographic to a set of concrete elements on which the pornographic character depends. We can simply say that if we were to remove some element of the object -- the background of a photo, say -- the picture would still be pornographic. If we continue to "feed" a recursive function the modified photo, each time removing some article which is not necessary to the pornographic character, we may eventually cross the line which separates pornographic photos from all other photos. But even then, we will not be able to say with certainty that pornography reduces to the status of the photo immediately before it crossed the line. The stripped down version of a photo may be accidentally rather than essentially pornographic. Moreover, a wholly different object may be equally pornographic. Similarly with the "active" supervenience manifested by Clausewitz's friction and Walzer's supreme emergency: a recursive algorithm is especially well-suited for tearing-down and building-up the supervenient scenario, although in the end it is not clear that we will ever reduce the concepts to a well-defined set of concrete circumstances.

## 4. Weber's (1996) analysis: four questions

Weber (1996) offers a stimulating discussion of the reducibility of fitness and of what it means to say that fitness supervenes on physical properties. He takes his initial direction from Kim's discussions of supervenience (Kim 1978, 1984, 1990[20]), particularly from Kim's distinction between covariation and causal dependence. "I shall argue that supervenience theories as accounts of the relation between fitness and physical properties are incomplete, because supervenience fails to distinguish between covariation and causal dependence" (Weber 1996: 411). Despite his unequivocal title, "Fitness made Physical," Weber begins and ends circumspectly, predicting at the outset that his overall argument "moves the picture closer to a reductive materialist theory of biological properties" (ibid.: 412) and concluding at the end that his approach "brings fitness and the theory of natural selection closer to the laws of physics and chemistry and to the generalizations of physiology, biochemistry, and ecophysiology than most writers in the philosophy of biology have previously thought" (ibid.: 429). It is significant that he stops short of claiming he has *proved* the possibility of a thoroughgoing material reductionism of any extent, even a "micro" one. A key phrase in Weber's rhetorical vocabulary in this paper is *multiple realizability* (as defined by Rosenberg 1978, 1985), a phenomenon demonstrated when two organisms in the same environment possess the same fitness despite having different physical properties. Weber accepts Rosenberg's basic contention here (Weber 1996: 412). What bothers him is Rosenberg's "claim that fitness *supervenes* on physical properties. This proposal appeals (says Weber) to the physicalist intuition that two things that are indiscernible with respect to their physical properties cannot differ with respect to any other property ('no difference without a physical difference')" (Weber 1996: 412-413). In fact, however, Rosenberg's use of supervenience seems at worst ambivalent with respect to physicalism as Weber seems to define it. Indeed supervenience à la Rosenberg may tend toward an anti-physicalism. Weber's argument may be read in different ways, depending on how one answers the questions below.

### (1) A first question about Weber's analysis: What does he mean by "explanation"?

Weber suggests three possible ways in which fitness and physical properties can co-vary: (1) Fitness and physical properties can be causally related to one another, that is, fitness can vary as a function of physical properties; (2) fitness and physical properties can vary together as functions of another, independent motive force; (3) fitness and physical properties can co-vary randomly (by chance). Following Kim, Weber rightly notes that calling F supervenient on physical properties does not tell us which of the relations (1) - (3) above applies. He makes it clear that he intends to argue that relation (1) is in fact the case. But oddly, he seems to suggest that if possibility (2) or (3) applied, then fitness could not be "explained":

> "This is not just a semantic problem. Whether fitness values merely co-vary with physical properties, or whether fitness causally depends on the latter is a crucial issue with respect to the question of whether fitness differences among organisms can be *explained*. If there was a mere co-variation instead of a causal dependence between fitness levels and physical properties, the situation would indeed be as hopeless as Rosenberg thinks" (Weber 1996: 413).

Two things are unclear here. First, what does Weber mean by "hopeless" as applied to Rosenberg's outlook? Perhaps he has in mind the general claim that biological science is theoretically but not practically reducible to propositions about more basic physical properties. But in fact Rosenberg's thesis can be read as being compatible with Weber's microreductionism, a point we will return to below. Or maybe Weber is thinking of his interpretation of Rosenberg's use of supervenience as physicalism. Let us accept uncertainty here and turn to a second matter in need of clarification: What does Weber mean by "explanation" in the quotation above? "Mere co-variation" in so far as it is opposed to causal dependence of fitness on physical properties must refer to sense (2) or (3) of co-variation. But presumably if we could trace the variation between fitness and physical properties to an independent cause, then that would constitute an explanation of the co-variation (contrary to Weber's assertion). Even a case of chance co-variation might be explainable in some sense if we could prove and characterize the randomness by statistical analysis or if we could demonstrate that the cause of variation in fitness were independent of the cause of change in physical properties.

Later in the same paragraph Weber uses the phrase "*causal* explanation" (my emphasis), by which he apparently means an explanation based on a causal link between fitness and physical properties. This is of no help in our attempt to understand what he means by explanation in general, however. A thesis such as, "Only in a case where fitness varies in causal dependence on physical properties can the relationship be explained as a causal function of physical properties" would be uninteresting. On the other hand (and to repeat), if Weber really intends to defend the thesis that "Only in a case where fitness varies in causal dependence on physical properties can the relationship be explained, period," then his understanding of "explanation" remains unclear.

### (2) A second question about Weber's analysis: In what way does he consider fitness to be analogous to other physical properties?

A second question about Weber's argument stems from the fact that he appears to draw the following analogy:

electrostatic forces:liquidity
::
various physical properties:fitness in a given environment.

Still more generally, Weber wants to argue:

physical forces: their manifestations
::
various physical properties:fitness in a given environment

But Weber ignores what seems to me to be a fundamental problem with these analogies: electrostatic forces and some other physical forces are defined and definable on a deterministic rather than a stochastic substrate. Under laboratory conditions, for instance, we can control electrostatic forces in such a way that a certain liquidity can be predicted as a necessary result of the given forces. By measures of fitness depending on the outcome of sexual reproduction, on the other hand, there is always a stochastic element present, even in idealized descriptions. (I cannot prove that this element of uncertainty is not epistemological, but neither can it be proved that the indeterminacy is ontological. Neither am I certain that the stochastic element does not exist in real manifestations of liquidity.) The case is reminiscent of the numerical analyst's joke recounted above. The quantity two in an absolute sense bears a certain relationship to the quantity five in an absolute sense. But the quantity "two" which occurs after execution of several hundred thousand computer instructions may not bear that same relationship to a similarly generated version of the quantity "five." Thus we cannot say that $2_a{:}5_a :: 2_c{:}5_c$. In particular, our ability to predict some future relationship of the computer's "two" and "five" is much restricted compared to our ability to predict the future relationship of two and five in an absolute sense. This analogy is not perfect, but it serves to indicate a point which Weber might have clarified. The same general question applies to Weber's use of "energy efficiency" as a physical property "of the same generality" as fitness.

### (3) A third question about Weber's analysis: How does he intend us to take his analogy

One possible distillation of Weber's overall argument looks like this: Fitness is analogous to qualities of physical systems which we accept as reducible. Therefore, fitness itself is reducible (or microreducible, depending on how one reads Weber). Before tackling this argument we should accept that one man's analogy is another man's puzzlement. For an analogy to "work," we have to perceive the similarity linking two avowedly different things as overwhelming the difference separating them. Relevant to the distilled argument above, Weber makes a good and useful point: a physical relationship may be ever so complex but nonetheless exist. (I take it that Rosenberg 1985 makes the same point in a different context when he defends the theoretical reducibility but practical irreducibility of biology to more "basic" physical sciences.) What Weber does not do is recognize (let alone discuss) a difference separating the two "sides" of the analogy. Although it may be difficult to reduce electrical resistance to its component factors, for instance, resistance is nevertheless always *abstractly* definable as a relation between such factors independent of other environmental factors. Certainly factors external to the definition can be relevant in real-life scenarios, but at least an abstract definition can be framed which excludes many such possibilities as extraneous. By contrast, it seems that fitness is either definable in a wholly formal way which precludes any reduction to concrete circumstances (reduction presupposes some bridge between genera), or else a definition which takes account of concrete circumstances in some way must be hugely complex and perhaps so arbitrary that it offends the intuition which made it interesting in the first place. Recall that it is not clear what constitutes reproductive success. Among the reasons for this is that large and frequent litters do not equate with survival of offspring in the second and succeeding generations (Morris 1986: 174, 178).

### (4) A fourth question about Weber's analysis: How are we to exclude the middle between macroreductions and microreductions?

Weber speaks of microreductions, apparently following a tradition which he feels is already so well-grounded in the literature that he has no obligation to explain the concept. In particular, he seems to sense no obligation to tell us how a microreduction differs from reduction on a larger scale. The same tension is evident in Kim's treatment of supervenience. After briefly discussing the supervenience of pain on physiological states, Kim (1978: 154-155) suggests we

> ...consider tables: there aren't precise laws about tables and there isn't some uniform microstructure underlying tables. One might say, parroting the functional-state theorists, that tablehood can be "physically realized" in so many different ways that it is highly unlikely, empirically, that some physical state will be found that invariably correlates with it. But this isn't to say that tables are not microphysical structures; nor is it to imply that the properties of *particular* tables aren't micro-reducible. What

> then lies behind these convictions? Why do we think that pains, although we do not know any physical state correlating with them (in fact we may doubt that there *is* such a state), are determined by the microphysical state of organisms, and that tables, too, are determined by physical properties of objects?

When Kim speaks of "microstructure," we can take it that he means nothing more than a level of the table which is not visible (not macroscopic). He explicitly classes tables as "middle-sized objects" (ibid. 151). We can read Kim, then, as claiming (1) there are no true general propositions of the form "When elements A, B, ... are present and when they are combined in ways X, Y, Z..., a table results at the level of middle-sized objects." However, Kim seems to argue that (2) there is at least one such proposition applied to some specific table. What remains a riddle by this account is whether we can speak at a level higher than the proposition referred to in (2), even if not at the level of absolute generality referred to in (1). If so, the dividing line between micro-, macro- and plain-vanilla reductions is equally mysterious.

# Part Two: Approaches to Chance and Circularity

# Chapter Five: Propensity as a Defense against Chance

The propensity interpretation of fitness has two major goals: neutralizing chance and avoiding circularity. The first goal (the subject of this chapter) derives from the tension between our intuitive sense that the fitness level of certain organisms is belied by their actual performance -- measured in terms of longevity and reproductive success. This "gut feel" is particularly intense when performance seems to be dramatically hampered or furthered by random events. As for the second goal (treated in chapter seven below), "circularity" appears frequently as an allegation leveled against conventional accounts of fitness, but elsewhere "tautology" and "analyticity" take its place. The charges encapsulated in these terms need to be investigated separately. We will begin, however, by examining the first goal of the propensity interpretation -- neutralizing chance -- paying particular attention to the strengths and weaknesses of this interpretation in each of its pursuits.

## 1. Neutralizing Chance.

It is clear that chance occurrences can wreak havoc on our judgments of the respective levels of fitness among organisms of a taxon in cases where fitness is evaluated by comparing the respective numbers of offspring actually left behind by two or more individuals. What actuaries might call "acts of God" can fell what seem clearly to be the fittest organisms before they can reproduce, or can hinder reproduction so substantially that the fittest seem less fecund than their presumably less fit counterparts. Such acts are unusual occurrences which, though perhaps foreseeable as a class, cannot be predicted as individual, specific instances. We know as a general fact, for instance, that lightning and meteor strikes do occur and that they can be lethal, but we cannot predict which organisms will fall victim nor when. Furthermore, we can acknowledge the possibility that an organism having the best (based on criteria to be discussed) of each attribute important to its taxon in a given environment -- and what is therefore the organism which would live longest and produce the most offspring among its peers *ceteris paribus* -- may fall victim to such chance phenomena early in life. We will return to the issue surrounding foreseeability of general versus particular effects below.

Mills and Beatty (1979) offer three hypothetical examples of how chance events can confound theories of fitness which are based on actual numbers of offspring produced.[21] First, if two identical twins are struck by lightning, so that the charred twin leaves no offspring while the survivor does, a theory of fitness defined in terms of actual numbers of offspring would force one to conclude that the survivor was fitter than its genetically identical twin. Such a conclusion seems unwarranted, again based on an intuitive feel for what fitness means, and that is *prima facie* evidence that a definition involving *actual* longevity and *actual* numbers of offspring produced is misguided. Second, Mills and Beatty imagine two butterflies which are phenotypically identical in every respect except that one is camouflaged in such a way as to hide it from its principal predator while the other butterfly is not camouflaged. If the non-camouflaged butterfly nevertheless survives to leave offspring while the camouflaged individual does not, a concept of fitness based on actual offspring produced would again force us into drawing an intuitively unacceptable conclusion. Finally, the authors note that “an earthquake or forest fire may destroy individuals irrespective of any traits they possess” (1979: 268).

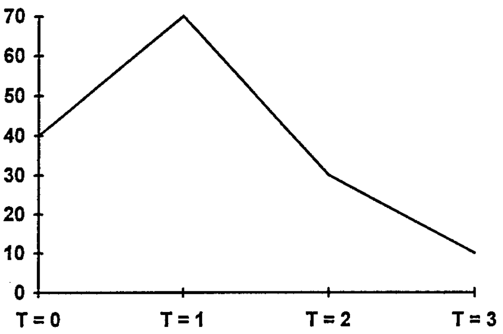
The propensity interpretation seeks to rescue our pronouncements on fitness from the vagaries of such chance phenomena by attributing a *tendency* to greater reproductive success even though the tendency may not be realized in fact. In Mills and Beatty’s version, this thesis can be summed up in a single sentence: “...the confusion [between fitness and actual descendant contribution] involves a misidentification of the *post facto* survival and reproductive success of an organism with the *ability* of an organism to survive and reproduce. We believe that fitness refers to the ability” (1979: 270). This statement seems to have been anticipated by Brandon’s criterion of “independence” of actual longevity or reproductive performance for a candidate law of nature: “If *a* is better adapted than *b* in environment *E*, then (probably) *a* will have more (sufficiently) offspring than *b* in *E*" (1978 in Sober 1984c: 64-66). Of course it is easy to create examples which make this seem a reasonable distinction. The strongest, fastest, most keen-sighted of a fish species, sprung from a venerable line of abnormally long-lived and extra fertile parents, may be vaporized when a satellite, its orbit decayed, splashes into the ocean. The young Überfisch thus had no opportunity to mate, but we feel certain that it had a propensity toward great fertility, or in other words that this particular fish was highly fit. Therefore we can properly call the Überfisch fitter than its peers of less spectacular construction.

It seems clear that the propensity interpretation succeeds here on a common-sense basis, at least to some degree. There are good, empirically defensible grounds for asserting that a given individual will be or would have been (had not chance intervened) fitter than its peers in terms of reproductive success. Breeders of horses, for instance, can predict with admirable accuracy not just which lineages are likely to produce qualities such as speed, endurance, size, etc., but also which are likely to be more or less fertile based on factors such as pelvic girth among females and the overall reproductive history of a given lineage. The most obvious example of this predictive ability is the case of the normally sterile hybrid, the mule; we can predict with a high degree of certainty that a given mule will be sterile.[22]

Yet because evolution by natural selection must in many cases work gradually, favoring individuals based on sometimes very subtle differences in comparison with their peers, we can justly ask of an explanatory device such as the propensity theory that it transcend common-sense, "broad-brush" distinctions and provide rigorous criteria for distinguishing between individuals. By this standard, the propensity interpretation seems vulnerable to the following interrelated problems when it claims to neutralize the confounding effect which chance has on our theories of fitness.

### (1) How a Propensity can Fail

Sometimes an organism leaves more progeny than another, although the one which is more fertile in fact seems to be less well adapted. As we have seen, a proposed explanation for this unexpected result is to appeal to the concept of propensity. Based on a propensity interpretation one can claim that an organism without progeny is fitter (better adapted) than one which has left behind many descendants. The key rhetorical move is to call the organism with no descendants fitter by finding a basis to claim that it had an innate *tendency* to greater fitness. One then proceeds to explain the dismal actuality by saying that the tendency, although real, was unrealized. An assumption of such an argument is that while an organism can fail to leave behind a given number of descendants, it cannot fail to have the propensity to a certainty degree of fitness. From this perspective, the tendency is a quality of the organism itself. In other words, the propensity to a level of fitness (i.e., a propensity to actual performance -- to a certain longevity or degree of reproductive success in a given environment, for instance) is a metaphysical fact, a basic property of the organism. If we choose to identify this propensity to a level of fitness as fitness itself, then fitness becomes a property of the organism, and most importantly, a property which is not vulnerable to random agencies. At any rate those are the outlines of the propensity argument.



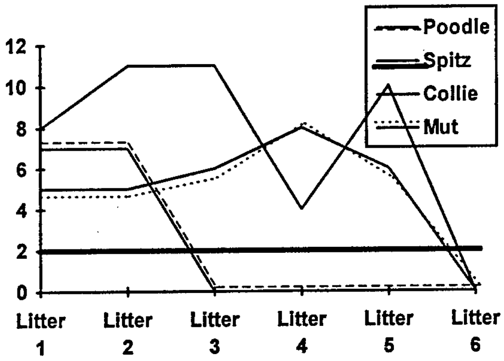
But we should  be skeptical of this claim. One can distinguish between two kinds of propensity. One sort is a historical tendency, that is, a general description of the direction of a given motion viewed from a necessarily limited historical perspective. In this sense we can say, with time T = 1 as our vantage point, that the curve in the figure above tends to climb. But it would be false to say in a second sense -- a general, non-historical one -- that the curve climbs, since we do not know what happens when T > 1.

Beatty and Finsen (née Mills) demonstrate sensitivity to this and related problems in their critique of and expansion on their own propensity interpretation of fitness (1989: esp. 19 - 29).[23] But the authors seem never to question the validity of a propensity interpretation so long as it is understood that the key term -- propensity -- can refer to any of a locus of well-known functions associated with probability distributions rather than merely to a single function.

> ...[T]here are many statistics on a probability distribution: the expected value, the mode, the median, the variance, the skew, etc. Why reduce fitness distributions to just one of these statistics? (ibid.: 23)
>
> The various statistics on fitness distributions, such as the geometric or arithmetic mean, the variance and the skew, are aspects of an organism's or type's reproductive strategy, and as such are components of fitness -- they contribute to evolutionary success in different ways, depending upon the environmental (broadly construed) circumstances. (ibid: 28 - 29)

Thus the challenge exercising Beatty and Finsen in 1989 is not to distinguish between a "reality" versus a propensity interpretation of fitness (as it had been ten years earlier), but rather to decide what kind of propensity fitness is. For instance, is it geometric or arithmetic? Is it to be measured in terms of descendant contribution to the next or to the 100th generation, reckoning from a given point?

What Beatty and Finsen do not emphasize is the possible failure of whatever sort of propensity interpretation a given researcher might have adopted. The example represented by the graph above demonstrates that, epistemologically speaking, propensities can fail to be ontologically real just as actual production of offspring can fail to occur. The odd phrase "fail to be real" should be understood as differing from the phrase, "fail to be realized." When a propensity to fitness is *not realized*, we take it that the tendency did not bear fruit in the form of offspring. But attribution of a propensity itself can fail to correspond to any reality. This may sound paradoxical at first, particularly if one understands by "propensity" an innate quality of an individual, a quality which either does or does not exist in that individual's overall makeup. But the realm of discourse here should be held constant and not shifted from an epistemological to an ontological focus. The propensity theory seeks to answer a question which arises in an epistemological vein: How can we compare individuals to judge their fitness without being confounded by chance? Or to put the question another way, What can we know of individuals (and how can we know it) that will allow us to speak meaningfully of fitness independently of chance occurrences? To reiterate, there is an unacceptable *epistemological* alternative: counting numbers of actual offspring is something we can do to gain one type of knowledge, but knowledge of this sort is sadly vulnerable to chance occurrences. The remedy, again an epistemological one, is to seek to know propensities rather than *post facto* numbers of offspring produced. But this prescription is also liable to failures of an epistemological nature. Perhaps a further example will help clarify the senses in which such failures are possible.



This chart shows numbers of offspring surviving to four weeks among four individual dogs. For instance, seven of the poodle's puppies in both the first and second litters survived to at least four weeks. It is obvious that evaluating tendencies here can be problematic. First there is the problem of *inferring a future trend.* Five weeks after the dogs' second litters, we might be tempted to infer that the poodle is fitter than the mutt. Hindsight would apparently contradict this premature conclusion. There is also potential difficulty in *interpreting past data.* The statement, "The collie tends to have fewer offspring after the fourth litter," is correct (i.e., the number of offspring from litters one through four is 34, for an average of 8.5 per litter, while the number of offspring from litters five through six is 10, or 5 per litter). However, the opposite statement, "The collie tends to have more offspring after the fourth litter," is also true in a sense: the number of offspring in litters five and six is 10, for an average of 5 pups per litter, compared with the 4 pups of litter number four. The salient issue is whether one chooses a point of a data "flow" and speaks of trends with respect to that point, or whether one compares two larger chunks of the flow with one another in order to formulate a trend.

In short, Mills and Beatty (1979) are right when they claim that we can be led astray if we seek to know the fitness of individuals based on the numbers of offspring which they actually produce. But the analysis of propensities is vulnerable to failures of knowledge as well, in part based on difficulties in interpretation and in part on problems of inference. Beatty and Finsen (1989) recognize that a single propensity interpretation (e.g., based on the arithmetic mean of a distribution) might not serve the purposes of all researchers all of the time, and might even mislead at times -- for instance, when copious short-term reproduction leads to a species' demise in the long run. How, then, do we characterize the functions which constitute the propensity interpretation of fitness? Beatty and Finsen conclude their 1989 article with the assertion that "'fitness' does seem to stand for a very broad family of propensities -- a family that is difficult to describe in general terms" (1989: 29).

I suggest that a recursive interpretation of fitness accommodates the necessary generality (though I realize that some may charge that such a reading is *too* general). In saying saying that the fitness of any genotype, individual or taxon at a given time is a function of that unit's fitness at an earlier time, very little restriction is placed on the function itself. Any of the calculations which Beatty and Finsen suggest could be associated with a propensity interpretation can be expressed by such a recursive function.

We continue with our analysis of problems facing the original propensity interpretation.

## (2) Ontological Propensity

If propensity conceived as an *ontological* property cannot "fail," strictly speaking, it is nonetheless a nebulous concept. It may be objected that by "propensity" we do indeed mean a real, ontological property of an individual, a quality which exists independently of our ability to know it. But even in an ontological sense, propensities are less robust than may at first be imagined. Suppose we claim that the Überfisch was fitter than its peers not because it produced more offspring (which it did not, having been vaporized so early in life), but rather because its known phenotypic qualities--e.g., its size, strength, and visual acuity--are demonstrably associated with reproductive success in other individuals, ones who did prove fitter than their peers in terms of actual offspring produced. This process of inference from many observed cases to arrive at a general, predictive statement ("Fast, strong fish are fitter," say) is our only means of establishing propensities in the first place.

But suppose it turns out that the Überfisch has such remarkable speed and strength because its hatching ground was contaminated by industrial waste of a certain type. Had the Überfisch not been vaporized at a tender age, it would have proved sterile. Or, if not sterile, it may have produced offspring suffering from a genetic mutation which would have caused them to be sterile. In this case, it is in fact false to call the Überfisch fitter than its peers who were hatched several miles away. Yet, the hardened propensity devotee will respond, there *is* a fact of the matter, even if we have misapprehended or ignored relevant facts. "But an object's *propensity* to manifest a certain property is a function of all the causally relevant features of the situation, independent of our knowledge or ignorance of these factors" (Mills and Beatty 1979: 273, n. 8). The case of the Überfisch again highlights an epistemological "failure," the propensity theorist will maintain, rather than an example which threatens the viability of the propensity interpretation in an ontological vein. Actually, however, propensities are at the mercy of chance just as actual reproduction is, for an unforeseen change in environment can erase a real (ontological) propensity to fitness. To avoid this danger, propensity theorists take pains to add the caveat "in a given environment" to their statements about fitness (or some similar caveat such as Coffa's "extremal clause," referenced by Mills and Beatty 1979: 278). In other words, attribution of propensities make sense only if couched in terms of conditionals: *if* the fitness environment remains the same, *then* organism x has a propensity to a certain level of fitness. There is nothing wrong with such cautious phrasing. In fact, it seems well advised. But it should be clear that a chance alteration in environment can obviate not just real reproduction and therefore "non-propensity" fitness, but also propensities themselves and even when they are conceived as being metaphysically real. In that sense, the propensity interpretation of fitness does not perfectly fulfill its self-appointed mission of escaping the ravages of chance.

### (3) Higher-order Fallacy

Another weakness of propensities in this context is what we might call "higher-order fallacy." The fallacy can be stated, admittedly with a very Aristotelian flavor, as follows: What an object in motion *tends* to do is a more accurate descriptor of the object's essence than what the object actually does. The name "higher-order fallacy" provides an interesting *double entendre* in this context. First there is the notion that a trend is more informative as to an object's essence than a particular fact. In other words, the propensity is a "higher order" of meaning. Secondly, the phrase "higher order" is commonly used when discussing derivatives, so that $d^2y/dx^2$ is said to be of higher order than dy/dx. But note the danger here. Let us assume that dy/dx measures speed, so that $d^2y/dx^2$ measures acceleration. It is possible that dy/dx will be greater at a given point on a curve describing motion than it was at a previous point (that is, the speed is increasing) while $d^2y/dx^2$ will fall between the same two points (the rate of increase is slowing). An analog in population biology might be the case in which population is growing, but at an ever-slowing rate. In such a case, what is the "propensity" of the population? Is it toward growth or stasis? Mills and Beatty apparently never considered such a case in which propensities applicable to a single scenario are heterogenous across "orders" (i.e., where a lower-order propensity demonstrates one trend while a higher-order propensity indicates a different tendency). The fact is that in observing motion of any sort, we may choose to search for the trend, the propensity, evidenced by the data. By that we mean what in differential calculus is called a higher-order derivative. But is a higher-order trend in any sense truer of the object in motion than a lower-order movement? A propensity is, after all, just another kind of movement and perhaps is as *accidental*, again in Aristotelian terminology, as a lower-order movement. Further, the trend or propensity itself has a trend or propensity, too, just as $d^3y/dx^3$ is meaningful despite the fact that our language lacks the neat, one-word expressions we have for dy/dx ("speed") and $d^2y/dx^2$ ("acceleration"). In short, the notion of a propensity is ill-defined, thus we cannot be sure that a higher-order propensity is a more chance-resistant "thing" than that of which it is a tendency, nor can we be certain that layers of propensity will not contradict one another.

### (4) The lack of a rigorous definition for "chance" occurrences

A further problem with proposing that a propensity interpretation can neutralize the bewildering, theory-confounding agency of chance is that it is not clear what constitutes the norm of an adaptive environment, and what a chance or "freak" occurrence means as a sort of complement to the norm in the range of possibilities. The case in which the Überfisch gets vaporized by a blazing satellite as it hurtles into the ocean seems clear-cut: satellites don't plunge through Earth's atmosphere every day, nor even every year, and when they do, a particular fish is not likely to be in the wrong place at the wrong time. In short, the specific occurrence is a rarity. But what about diseases which affect the fish population? If an infectious disease *regularly* decimates the species every century or so, and the Überfisch happens to have weaker immunity to that particular malady than its supposedly less fit peers, do we call the disease a *chance* occurrence and continue to maintain that the Überfisch is fittest? What does "rarely occurring" mean in this context? One might attempt to argue that if a disease ravages a species only intermittently, say every fifteenth generation, that the adaptive environment of the species changes in those generations. However, such a rhetorical maneuver eviscerates the spirit of our inquiry. If we lack a stable standard for deciding what is a normal occurrence and what a fluke, we lose any chance of making clear the distinction between the creation of offspring and the tendency to create offspring.

Let us approach the matter from a slightly different angle. Sometimes theories about how conceptual entities correspond to one another are motivated by the desire to avoid paradox. Conceptual confusion can occur when a key theoretical commitment is threatened by another aspect of the theory. Dawkins' discussion of the selfish gene (1976), for instance, is prompted largely by his desire to show that altruism at the level of individual organisms can be explained by taking the gene as the unit of selection. That perspective (claims Dawkins) allows us to observe apparently selfless behavior of individual organisms without abandoning the belief that whatever units are ultimately affected by natural selection struggle against one another in a wholly selfish fashion (metaphorically speaking, of course). Whether one believes that Darwinian evolution requires the commitment to selfish struggle, as Dawkins apparently does, is a matter to be considered later. For our present purposes, what is important here is that the concept of the selfish gene is largely a means of avoiding the paradox which individual altruism presents, given the commitment to selfishness.

A similar motivational pattern can be seen in the case of the propensity interpretation of fitness. First, the conceptual hazard, the paradox, is set up. Then the authors rework what they take to be the aspect of evolutionary theory which allows the paradox to creep in. This aspect turns out to be what we are calling a "reality" interpretation of fitness. Here is one of Mills and Beatty's examples illustrating the paradox:

> Scriven (1959) invites us to imagine a case in which two identical twins are standing together in the forest. As it happens, one of them is struck by lightning, and the other is spared. The latter goes on to reproduce while the former leaves no offspring. Surely in this case there is no difference between the two organisms which accounts for their difference in reproductive success. Yet on the traditional [actuality] definition of "fitness," the lucky twin is *far* fitter. Most undesirably, such a definition commits us to calling the intuitively less fit of two organisms the fitter if it happens that this organism leaves the greater number of offspring of the two (1979: 267; reprinted Sober 1984b: 267).

This identification of the undesirable paradox by means of examples brings Mills and Beatty to a crossroads. They can choose to combat the paradox directly, by removing the conditions which cause it. They seem on the verge of attempting this when they state their rhetorical goal as follows: "In general, we want to rule out the occurrence of any environmental conditions which separate successful from unsuccessful reproducers without regard to physical differences between them" (1977: 272; Sober 1984b: 44). But in fact Mills and Beatty never attempt to "rule out the occurrence" of such "environmental conditions," if by that they mean, first, proposing a rigorous criterion for identifying what constitutes such environmental conditions and, second, suggesting that we ignore reproductive phenomena occurring in the shadow of such conditions. Let us take these issues one at a time.

First, it should be made clear that Mills and Beatty never offer a rigorous and general account of what they see as the ultimate source of the problem with "reality" interpretations of fitness, namely these "environmental conditions which separate successful from unsuccessful reproducers without regard to physical differences between them" (ibid.). Instead, the reader is left to infer the problem from the examples, while foundational questions go unasked and therefore unanswered, hidden behind unstated assumptions.

In a *prima facie* way I find the lightning example rather convincing, as I assume most readers will. Further discussion will therefore try some average readers' patience, but perhaps it will prove worthwhile to proceed with an anatomy of the example nonetheless. The most striking assumption in the lightning example is that the bolt's effect on reproductive data is deceptive rather than enlightening. It is stipulated that the twins are standing together, but if they were very near, then both would have been equally affected by the bolt. Since they were standing in different places, how can we rule out the possibility that a behavioral difference (identical twins are not, after all, truly identical in either behavior or structure) caused the deceased twin to perish and allowed the other to survive? At this point some readers may cry foul, insisting that the example deserves a charitable reading and that it is, after all, rather convincing. Empirically speaking, lightning seems to strike organisms dead unpredictably, without regard to their fitness. This observation may be true, but is it sufficient to allow the example to perform as Mills and Beatty intended? We know, for instance, that among *Homo sapiens sapiens* golfers have a rather high incidence of mortality due to lightning strikes as compared with the rest of the population, yet each person who dies on the links, viewed individually, is apparently the victim of an unpredictable phenomenon.

The easy case is one in which there is a marked difference between two otherwise very similar organisms, such as twins. Let's say the difference in this case is behavioral and applies to human twins: one twin likes to golf and the other likes to watch TV. We can assume without further argument that the couch potato twin is less likely to be struck by lightning than his links-addicted brother. (What this means in terms of overall fitness is unclear, since the twins' environments cannot be exhaustively defined in terms of TV watching versus exposure to electrical storms.)

Now suppose the couch potato happens to flip to a sports report. Inspired, he leaves his living room and begins to play soccer four hours every day. A quick look at actuarial tables covering data gathered in the US shows that the soccer player is much less likely to be struck and killed by lightning than his golf-playing twin, so the behavioral change still leaves one twin more vulnerable than the other based on factores which are not describable as wholly random. The example can be altered again such that both twins play golf, on the same days, at the same time of day, etc. The point would be to force the example to show what Mills and Beatty want to demonstrate -- that at some point there is no difference internal to the two individuals which explains why one is killed and therefore leaves no offspring while the other finishes his round, showers in the clubhouse, then goes home and impregnates his wife. Nor (if we construct the example carefully) is there a conceivably significant difference in the twins' environments: each stands in a slightly different spot at the moment the fatal bolt strikes, but there is no reason to think that the difference in location is significant if we rule out the victim's possibly nearer proximity to water and to the pin on the green, and similar factors. Even if we allow such environmental aspects to play a role, there must be some stipulation that more than chance made one twin avoid the danger areas or seek out the relatively safe zones before we can attribute the death either to differences between the twins or to differences between their respective environments. That leaves us with just one explanation of the event -- chance. And from this Mills and Beatty conclude that a good analysis of fitness cannot be based on looking at real differences in reproductive rates. Any single event might mislead us by resulting from chance instead of a difference between two organisms. Instead, so the theory goes, we must look at tendencies.

Two responses seem obvious but are not treated explicitly by Mills and Beatty. First, as soon as we start looking at more than one instance of differential fitness, we stand a good chance of seeing differences in which the workings of chance cannot be ruled out. Thus we will *necessarily* look at tendencies, which is to say that a propensity interpretation of fitness is unavoidable. In the reality of a biologist's work, there is no such thing as a reality interpretation when confronted with the question, "What, exactly, is the fitness difference between organisms possessing qualities a, b and c, and those possessing qualities x, y and z?" Secondly, in cases where the scope of observation or a paucity of data leaves us no differences among what we take to be identical cases, there is no question of looking for propensities. Reality is all we have.

If the fitness of an organism depends upon its environment, then fitness can be a stable quality of the organism only to the extent that the environment is stable. But what does this mean, for the environment to be *stable*? An observer recognizes stability through its *predictability*. If some event occurs which is not predictable -- which happens (to use Mills and Beatty's favored phrase) by chance -- then they implicitly exclude it from the selective environment *per se*. If it were a part of the environment, then it could tell the observer something about an organism's fitness rather than distorting any evaluation. Perhaps this point will be clearer if we consider two further scenarios offered by Mills and Beatty.

> The counter-intuitiveness of the traditional definition is also suggested by the following hypothetical case. Imagine two butterflies of the same species, which are phenotypically identical except that one (C) has color markings which camouflage it from its species' chief predator, while the second (N) does not have such markings and is hence more conspicuous. If N nevertheless happens to leave more offspring than C, we are committed on the definition of fitness under consideration [the actuality interpretation] to conclude that (1) both butterflies had the same degree of fitness before reaching maturity (i.e., zero fitness) and (2) in the end, N is fitter, since it left more offspring than C. (1979: 278; reprinted in Sober 1984c: 40-41).

The propensity interpretation is similarly motivated by the intent to avoid paradox, or more precisely, to escape a range of conclusions which would not jibe with certain key theoretical commitments. First, there is what we might call the "unequal twins paradox," which Mills and Beatty illustrate with a scenario borrowed from an earlier work.

> "Nor can these counterintuitive results be avoided by shifting the reference of fitness from individual organisms to groups. For, precisely as was the case with individuals, the intuitive[ly] less fit subgroup of a population may by chance come to predominate. For example, an earthquake or forest fire may destroy individuals irrespective of any traits they possess. In such a case we do not wish to be committed to attributing the highest fitness values to whichever subgroup is left" (ibid.).

This is the clearest formulation of the paradox Mills and Beatty wish to avoid because the commitment to a trait-based concept of fitness is most explicit in this statement. The key phrase, "irrespective of any traits they possess," is also the most controversial aspect of the example. Inane though the question may seem at first, we should nevertheless pose it: How would we know that an organism destroyed in a fire is not less fit than other organisms which survive? Of course we can formulate hypothetical examples in which it is assumed that survival or death is a coincidence. But it might be quite difficult to prove that some phenotypic difference among individuals does not explain the early mortality of some and the longevity of others.

In the final analysis I think Mills and Beatty's development of the propensity interpretation is helpful, although the issues raised in this section are worthy of consideration nonetheless. Section 2 below (under the heading "Can we defend an 'actuality (or reality) interpretation' of fitness?") raises further questions about the range of scenarios to which the propensity interpretation can profitably apply.

### (5) The chance element of sexual reproduction and its effect on a propensity interpretation

Here, once again, is a key passage indicating what Mills and Beatty have in mind when they say that fitness is a propensity: "But an object's *propensity* to manifest a certain property is a function of all the causally relevant features of the situation, independent of our knowledge or ignorance of these factors" (Mills and Beatty 1979: 273, n. 8). The authors mean that fitness as a propensity is indeed an ontologically real property of a given individual. The only confusion admitted is with respect to how one reckons the numerical value of this property; "expected value" is proposed as the proper interpretation.

But to remove the chance element from the equation may be impossible. To see this, let us consider the case of organisms which reproduce sexually. If we say that organism A has fitness value $f_1$, calculated as a function of the expected number of its offspring, we must somehow take into account the fitness of A's mates as well. And how are we to know which mates are to be considered, or indeed which mates are available? Fisher's 1930 hypothesis that the average number of offspring is greater for the rarer gender in a population suggests that the gender of A's contemporaries is crucial to the calculation of A's expected number of offspring and therefore of A's fitness. Yet the ratio of genders may be in constant flux. Are we to take the caveat "in a given fitness environment" to include up-to-the-minute information on this ratio? Even if so, we cannot know how many of A's offspring will be alive until we know the genetic composition of each of A's possible mates, as well as a means of weighting each organism's probability of becoming A's mates. Such complications (and we could certainly list more of them) argue for a recursive interpretation of fitness. As we have seen, the number of iterations which a recursive algorithm will spin out is formally unlimited, which means we can manipulate a function defining fitness to take account of an unlimited number of such complicating factors. In short, the dynamic character of the possible scenarios demands a formal definition which is more dynamic than a straight, non-recursive propensity interpretation.

### (6) Summary: Foreseeability of general effects versus the foreseeability of specific effects

What do we mean when we say that we know lightning will kill some individuals, but that its agency cannot be taken into account except by saying that it is a random or chance occurrence? Presumably we mean that the following statement will be true, but that we cannot assign values to the variables, that is, that we cannot say which particular individuals will make up the set S nor which specific times the set T will comprise:

(C1): *The set of organisms S:{s1, s2,...,sn} will be killed by lightning strikes at respective times T:{t1, t2,...,tn}.*

Supposing that a naturalist were assigned to observe a species which she had never before observed, in an environment which was equally new to her, she might at best make a few tentative guesses as to *how many* organisms would be struck by lightning within a given period, based on her knowledge of other, somewhat similar organisms within a like environment. Later, after years of observation, the same naturalist might have formed inductive rules which would allow her to estimate with greater accuracy how many of the organisms under consideration would be killed. Perhaps she could even offer a few generalizations about which specific organisms would be more likely to be killed. For instance, maybe males tend to indulge in dominance displays and contests in open fields more frequently than females, making males the more frequent casualties. And it is conceivable that among males, those who open their eyes first among their litter mates tend to be more active in displays and contests than those of their siblings which develop more slowly. Thus the naturalist could point to a specific individual within a given litter and say with some assurance that that animal is *the most likely* to be struck by lightning. Other animals -- a litter mate who dies in the first week, for instance -- may be ruled out as future victims of lightning strikes. For such an educated naturalist, probability still figures very large in a statement such as C1 above, but at least there is more to be said about the phenomenon of death by lightning strike than that it is a randomly occurring possibility.

How does claim C1 differ from other claims about future events relevant to the fitness of organisms? Let us assume for the moment that all claims about the future of organisms are based on inductive generalizations from observation of past occurrences. For instance, we might wish to address the mating habits of certain animals within a given environment. A relevant question in this field of inquiry would be, Which individuals among all those born within a two-week period will mate successfully? Because the question deals with the future, absolute certainty may be impossible. In other words, an element of chance may inevitably remain. But by examining the organisms themselves, a negative kind of certainty may be attainable in some cases. Analysis of a newborn's genetic makeup, for example, may tell us that the individual is sterile. Apart from such negative cases, however, an element of chance remains, just as in the case of C1 above. An educated naturalist may be able to tell us many interesting probabilities, inferred from past observations, but all such information is essentially a statement of *what probably happens to individuals of a certain profile*. In this sense any statement about the future of organisms will have the form of C1; one can simply substitute a description of the phenomenon of particular interest for the phrase "will be killed by lightning."

This might be taken as an argument *for* the propensity interpretation. Because the functionally defined type to which an individual belongs is determined by properties of that individual, it might be argued, why not claim that all matters relating to the future of organisms -- fitness included -- are determined by propensity? (Here "functionally defined" means determined by the purpose at hand, so that a "type" might be something usual such as a taxon, but it could also be a category of organisms described by a phrase such as "will be killed by lightning.") Mills and Beatty do not go this far, perhaps because they do not interest themselves in the question of whether all statements about the future of organisms must be interpreted in the vein of a propensity. But it does seem clear that the authors propose their propensity interpretation as a means of circumventing the kind of paradox which chance can introduce -- calling an organism fitter than its identical twin because the twin was killed by lightning early on in life, for instance. What needs to be emphasized is that this strategy depends on assumptions which one might not wish to accept. From an epistemological standpoint, it is possible that, by chance, one misconstrues the relationship between type and the way that type generally fares. Perhaps a naturalist observes twenty generations of elk and based on the data collected forms a profile of the theoretically fittest individual. Clearly the data can be affected by chance; they can also be misinterpreted. Mills and Beatty suggest that fitness propensity is a real, ontological property of an organism. They might therefore claim that the possibility of an epistemological mistake does not affect the ontological reality, that is, that there is a fact of the matter with regard to propensity and that this fact of the matter is independent of perception (and misperception). The authors do not seem to take into account, however, the possibility that the interaction between environment and individual is not necessarily deterministic. Perhaps the sum total of all relevant features of environment and organism leaves open several possibilities, much as the two-hole experiment demonstrates a random facet of the universe at the particle/wave level. Such a non-deterministic interaction is not to be remedied by describing all parameters (Feynman 1965, pp. 127 - 148).

From the point of view of evolutionary biology, this line of speculation may seem to be a dead-end. Whether that is the case or not, it appears that the philosophy of evolutionary biology must at least seriously consider the possibility analogous to the two-hole experiment: perhaps there is a real element of indeterminacy at the level of biological evolution just as there is at the level of particle physics. If we cannot say in advance which organisms will survive to which age and which will have a given number of offspring, regardless of how completely we can describe the organisms and their theoretically stable environment, then the goal of establishing a theory of fitness which escapes all paradoxes will not be reached.

### 2. Can we defend an "actuality (reality) interpretation" of fitness in some cases?

Yes. That, in a word, is the thesis of this subsection. While Mills and Beatty are undoubtedly correct in arguing for an interpretation of fitness which is based on tendency rather than reality *in some circumstances* (we will soon examine what criteria define such circumstances), there is absolutely nothing amiss with adopting a reality-based interpretation of fitness in other situations. Which interpretation makes the most sense, reality or propensity, depends on three criteria -- the "temporal direction" (forward or backward) of an analysis, the sense in which the results of the analysis are quantified, and the organizational unit of life on which the analysis focuses. This dependency will be clarified by considering each of these three criteria in terms of distinctions between, respectively, *forward-* versus *backward-looking analyses*, *Boolean* versus *graduated quantification*, and *grouped* versus *stand-alone subjects*. (The meanings of these terms will be clarified shortly.) Moreover, the *actual* longevity and reproductive success of organisms *must* come into the picture at some point, else the notion of propensity can have no empirical meaning in a specific case. Empirical vacancy would in turn be deadly for a scientific theory such as the theory of evolution. But that observation jumps ahead of the argument, so let us begin by examining the three distinctions just mentioned.

### (1) Forward- versus backward-looking analyses.

The first distinction is between what we can call *forward-looking* versus *backward-looking* analyses. It seems clear that the notion of fitness, however it happens to be defined, is used in two broad, *temporally*-defined ways. First, we might ask about the future prospects of a specific organism. How long will it live? How many offspring will it produce and what will be their fates? This kind of forward-looking analysis need not be performed from the standpoint of a presently living organism. We might find a fossil and wonder what became of the lineage which it represented.[24] In the absence of further fossil data offering an explicit answer to our question, our analysis must begin from the standpoint of the organism which the fossil represents. Looking forward from that point, we can draw inferences from what we know about the organism's structure and about the environmental conditions which reigned from that time forward. In either case -- looking forward from the present or from some point in the past -- there is an indeterminacy, a lack of information as to what will happen in the future or what did happen in a past for which we lack data.

Such information gaps contrast with what we find in backward-looking analyses. Here, one can inquire into the fitness of organisms and taxa which have already lived out their reproductive lives or have demonstrated their reproductive capacities in some fashion. That is not to say that such individuals or groups are necessarily dead or extinct, but only to say that they have already provided some specific, indisputable data about their relationship to their environment and to their peers. When we see in the fossil record, for instance, that a certain group of organisms such as a species arose, reached its apex, then died out, we might well conclude that the final form of the species (assuming that there was some variation during its existence) was ill-suited to whatever environment happened to exist at the time of extinction. Of course there is the possibility that a rare cataclysm occurred, such as the meteor strike and resulting dust cloud which some have suggested could explain the near-contemporaneous extinctions of so many species of dinosaurs. It is further possible that we might not wish to consider such an apparently unusual event as being a part of the adaptive environment in the same way that more ubiquitous and regular features such as seasonal weather changes and average temperatures are components of environment. But in at least some cases of extinction, it is clear that a dearth of fitness was responsible. That this is a comparative judgment -- that is, that there is no absolute standard of fitness or "unfitness" in the case of a given individual or taxon -- makes no difference for our present purposes. What counts is that sometimes we can see a real phenomenon such as extinction as having occurred in the past, and from that *fait accompli* we can draw certain conclusions about fitness. (Of course there are cases in which we can adopt either a forward- or backward-looking perspective, as the purposes of research require.)

Now in some backward-looking cases it would seem superfluous to speak in terms of propensities rather than actualities. In the case of a fossil representing a now-extinct taxon, for instance, we *could* say that the taxon in question *tended* to be unfit, that it *had a propensity* to be ill-adapted to its environment, as compared with a peer species which continued to exist. But this is a senseless use of the notion of propensity or tendency. In such a case it is clear that the taxon in question *was* unfit as compared with other taxa which continued to exist.

In so saying we of course rule out chance on a monstrous scale, which should tell us something about chance as it functions in such discussions. Mills and Beatty rightly worry that a chance occurrence, such as a bolt of lightning striking down one organism but leaving its identical twin unharmed, destroys the validity of an actuality interpretation and drives us toward a propensity interpretation. (By a twist of fate, one of the twins has already outlived the other and stands to leave more offspring in actuality. Because the twins were identical, however, it is counter-intuitive to suggest that one twin is fitter than the other.) But an argument can always be made that any context, environmental or otherwise, is the result of chance. That in turn means that long-term environmental conditions are as much the product of chance as the lightning strike, so that large-scale, gradual extinctions allow us no more chance to talk of actual fitness than the lightning strike. Following this line of reasoning would force us to conclude that we can *never* speak of *actual* fitness, since chance always plays a role. That in turn means that we cannot even gather data for conclusions about propensity. There would be no context which is not chance-ridden to the extent that we would have to throw out data and look further.

Let me say immediately that I would not care to make such an argument. On the contrary, I think Mills and Beatty are correct to worry about counter-intuitive conceptions about fitness. However, such misconceptions arguably do not arise from observing chance occurrences, but rather by observing too few occurrences, and the two are not necessarily the same. As just mentioned, it is possible to consider our present climactic conditions to be a matter of chance. That does not mean just the weather as it happens to exist at the moment, nor the fact that this year may have been a particularly cool one as measured against the last three to four decades of meteorological data. Each of those aspects can be seen as the necessary result of high- and low-pressure patterns, jet stream locations, and any of a number of other factors. We can choose to view a given phenomenon as locally regular if we concentrate on the causal connection with such predecessor conditions. This group of causal associations forms a context of a kind, and so long as we do not go outside that "frame" to ask what other contexts might have originated, we tend not to see irregularities which we call random.

But at some point we must admit that things might have been different than they are, so far as we know. Perhaps our theoretical knowledge is so well developed, and our store of data so extensive, that we can trace a causal chain back thousands of years. In this chain, each occurrence is the necessary consequent of the occurrence before it. But at some point we will lack such certainty and conclude that things might have been different. One of the "beads" of the historical chain might have been other than it is, which in turn would cause all the other relationships between antecedents and consequents to have been different. What this means is that the purpose of a propensity interpretation is not to avoid being tricked by chance -- that we cannot accomplish -- but rather to *define* what we mean by chance, and that is done by increasing the number of observations and thereby *creating* a context of regularity. (Note the possible analyticity here: it appears that we define fitness as a propensity immune to chance circumstances as we simulatneously define those circumstances.) In fact we cannot say with certainty that the allegedly identical twins were identical (a mutation may have occurred), and there may have been something which made one twin more likely to be struck by lightning. If there is a gene which determines how likely it is that a person will become a golf player (an activity, played on exposed greens and fairways with metal clubs, in which many are struck by lightning each year), or will choose to live in a town such as Miami (in which lightning strikes are frequent) rather than, say, Helsinki, then there may be a genetic propensity to be struck by lightning as compared with other human beings. This is meant only half in jest. The fact is that we simply cannot say, except somewhat arbitrarily, what is and is not a matter of chance. To repeat, what we can do is to create conceptual chance-free zones by declaring what we understand to be the norms within certain observational contexts.

This point is important enough to warrant restating in a slightly different, perhaps more technical fashion. The axioms of probability theory depend upon the notion of a sample space. Thus the probability of something occurring plus the probability of the complement (of that thing not occurring) must cover all possible events...in the sample space. But here we need to be particularly careful in interpreting the word "complement." When we employ this concept as above we might mean a contrary, so that the probability of A, denoted P(A), has as complement the probability of not-A, P(~A). Thus with regard to a certain organism, there is the possibility that the organism is white, but there is also the possibility that it is not white. As in the case of all contraries, we cannot falsify both statements; that is, we cannot say that a given object is not white and not not-white. Or we might mean a *contradictory* when we talk about something belonging outside a given subset of the sample space. In such a case, it might in fact be true that an organism is not white and also not blue. The temptation is to interpret the first case as including all possibilities. In other words, we might assert that so long as we deal with contraries, we cannot possibly fail to take account of some possibility. But this would be false, since we might deal with a limited set of qualities in our limited sample space. Outside that realm, there is no telling what qualities we may encounter. We might have interested ourselves only in whiteness, whereas size may in fact be relevant to a judgement of fitness as well. (Consider the fact that while white serves to camouflage polar bears their fitness also derives in part from size, since large organisms maintain body heat more efficiently than small ones where large and small are reckoned in terms of the ratio of mass to surface area.)

What is important here is that the sample space itself is chosen, so that what constitutes regularity versus irregularity (or read chance) is also chosen -- created, we might say -- by the observer. What is rare and what is common depends, in short, upon one's frame of reference. Probably this will be clear to all readers without further argument, but perhaps a simple example is in order to make certain. It is clear that if one rolls a single, normal (i.e., six-sided) die, then the chance of the number six coming up is one in six, 1/6. But suppose that instead of being a cube, the die actually had twenty-four sides, numbered one in twenty-four. In such a case it is immediately clear that the sample space has changed, and thus the chance of rolling a six is one in twenty-four.

Let us apply these ideas to the discussion of forward- and backward-looking analyses. When we know that a taxon has become extinct, it is clear that the *real* longevity of individual members of the species and the *real* numbers of offspring they produced were insufficient within their environments. Even in the case of a "fading-out" term such as "fit," that means that the organisms, observed as a group, were *really* unfit in their selective environment whether that environment included random elements or not. Mills and Beatty's notion of propensity as a kind of *capacity* for various degrees of reproductive success simply need not come into play here unless we are convinced that some major element of chance occurred. But arguably we are less and less likely to judge chance to have been operative as the period of the now-extinct organism's reign is seen to be longer and longer. As the authors present their theory, its primary goal is to avoid paradoxes of the kind which arise when one twin is killed by a lightning strike which might just as well have killed either twin, both, or neither. In the case where one is killed in this random way, Mills and Beatty correctly observe that it is wrong to call the surviving twin more fit based on its greater longevity and possibility of greater reproductive success. Mills and Beatty are also right in considering the twins equally fit and in searching for a way to avoid a paradoxical conclusion. But this is only true given the assumption that the lightning strike was indeed a chance event. Thinking of fitness as a capacity for a certain degree of reproductive success rather than actual reproductive success counteracts the chance agency, or so Mills and Beatty argue. But what about the case where both twins survive to reproduce and then eventually die? Would it be misguided to compare their real reproductive records to that of organisms within their environment? The answer must be decidedly negative. In this *backward-looking* case, the real reproductive histories of organisms are an acceptable basis for drawing conclusions about fitness.

It is rather in the case of *forward-looking* analyses, where it is uncertain what degree of reproductive success an individual organism or species will or would have enjoyed, that the propensity interpretation must be preferred to a perspective of fitness based on actual case histories. It is important to repeat here that a forward-looking analysis can apply to cases in the past. For instance, an organism might appear in the fossil record with certain phenotypes clearly evident. However, we would probably not have any clear notion of what degree of reproductive success the individual represented actually experienced. We might then wish to speculate about the organism's probable fitness, perhaps based on the longevity of the type of organism possessing those observed features and a knowledge of the environment at the time or in other words of its probable reproductive success. This would equate with speaking about a *capacity* for reproduction, which is what Mills and Beatty mean by propensity.

There are three important points here to which we will return. First, in cases where we can talk about some real, verified phenomenon such as the extinction of one species versus the survival into the present day of another, we are generally justified in speaking of *real* fitness rather than just a propensity toward a certain degree of fitness. Secondly, when we must speak in a propensity vein for want of hard data or because chance events (meaning those of a certain rarity as judged by some non-objective standards) such as lightning strikes have taken a hand, our talk of propensities can only take on empirical meaning by induction from actual observed instances of longevity and reproduction. (The induction comes in with the correlation of reproductive data with phenotypes or with properties at some other level.) Finally, the judgment of which events are regular and which are the results of chance is a function of observational *cycles*. We *create* our conception of chance and regularity by choosing an observational context and then watching what occurs within that context over and over again. The sum total of *actual* results accumulated during these observational cycles is what allows us to speak of propensities.

In passing it is worth noting that the distinction between forward- and backward-looking analyses is especially important when we confront the claim that although we cannot say evolution is directed toward a specific goal in the sense that some have claimed the human species was evolution's goal and its crowning achievement (a point of view which Midgley 1985: 69 - 70 attributes to Lamarck), nor even toward a general goal such as complexity (cf. Gould 1994 for a criticism of this view), we can say that evolution by natural selection is teleological in a certain sense. A given environment will presumably have a set of constraints which, though extremely complex, define a set of *optimal* organisms. The optimum is of course relative, that is, determined by the environmental context, but there is nonetheless an optimal set of imaginable organisms. It does not even matter so much that we conceive this set as being small or even finite. What counts it that the set of optimally adapted organisms excludes many more kinds than it includes; that "narrowing" of structures indicates a teleology, a kind of goal-directedness. Nor does it matter that the mechanism (natural selection) produces different optima in different contexts rather than producing the same optimum universally. We may choose to think of environment as an essentially purposeless and unconscious entity, and then we may decide to believe that such an entity cannot be a designer and therefore that the production of optimal sets through the agency of natural selection is a non-teleological phenomenon. But neither belief -- that the environment taken as a whole despite its many changes is purposeless and unconscious nor that such entities (assuming the description is correct) cannot stand behind teleological phenomena -- is grounded in anything like a scientific theory. This is not to claim that the environment is conscious and purposive, but only to say that we do not have consensus definitions of these terms which would exclude the environment from the set of such beings. Nor do we know enough about the environment to make such a judgment even assuming that there were a scientific consensus as to what constitutes consciousness and purposiveness. And what if we were certain that the environment did not meet the standards which might someday be agreed upon? The notion of teleology, apart from any cultural overlays, does not require a purposive designer. In short, if we accept that natural selection tends, in a probabilistic sense, to modify existing life (at a level above the individual) so that it is better suited to the environment, then natural selection in a given environment can be viewed as a teleological process. (This conclusion seems consistent with Teilhard de Chardin 1956, 1957.)

The reason this is relevant to our present discussion is probably obvious, but it may be well to spell it out here nonetheless. In so far as chance and regularity are established by the selection and repeated observation of specific contexts out of the "soup" of all possible frames of observation in nature, there is virtually no way of avoiding a teleology. (Here the term does not carry the connotation of purposiveness.) This is because in so far as regularity emerges at all, it will imply to the observer that some set of conditions tends to result in the same set of outcomes. One may choose a context in which no such pattern emerges, but when the context chosen does display this kind of regularity, there is by definition a narrowing of possible outcomes given the same antecedent circumstances. That narrowing may be seen as a "local" teleology.

### (2) Boolean versus graduated quantifications of results

A second distinction compares what I will call "Boolean fitness" with "graduated fitness." It seems that we can talk in an all-or-nothing mode if we take our standard of analysis to rest on something like the difference between extinction and survival. When we look at a lineage of organisms in the fossil record, for instance, it may be clear to us that the lineage has either survived into the present day or else that it is now extinct, having either died out altogether or else been modified so dramatically that its present-day ancestors are judged to represent at least different species if not different genera or even higher-level criteria of classification. The concept of fitness can certainly figure in rough judgments based on survival and extinction: a species which survives longer than another can properly be called more fit (ceteris paribus). Such a distinction could be used to answer questions of the form, "Was species x fit enough to survive at least y million years?" The answer yielded by the fossil record will be Boolean -- either a yes or a no (though the caveat "assuming no examples to the contrary turn up" would have to be added after the negative answer).

On the other hand, it may be that some scientific purposes call for greater subtlety in comparing  organisms or taxa to one another. The question might be put, "How fit was the species within the space delimited by a certain latitude and longitude as compared to the same species elsewhere?" The answers "more fit" or "less fit" might suffice in such a case, but on the other hand, researchers might want to have actual numbers presented to them: a species' population in environment A increased four percent per year, while the same species in environment B only held steady. That description of the state of affairs would correspond to a "graduated" understanding of fitness.

Now at this point it might be objected that what we are calling a graduated concept of fitness is in some sense more basic than a Boolean understanding in that one can immediately draw a Boolean conclusion given graduated data, but not vice versa. For instance, if individual organisms have been assigned fitness values according to some scheme (as in Mills and Beatty's suggestion that fitness be understood as weighted probabilities that various numbers of offspring will be produced), then one can immediately judge whether a given organism is fitter than another. Given only the information that one species has survived and another has not, by contrast, we can say nothing with certainty about the individuals which made up that species. This may be true, but even if graduated standards of measurement possess some sort of primacy in this sense, that would not mean that graduated measurements are always possible. In fact -- as in many examples involving the fossil record -- a Boolean judgment of "more fit" or "less fit" is the best that we can do.

There may also be a theoretical objection to the abstract contention that Boolean standards can always be inferred from graduated standards, but not vice versa. To see this, it must first be realized that the fitness values which Mills and Beatty recommend do not conform to the actual numbers of offspring produced any more than the probability that a certain number of barges will pass under Heidelberg's Alte Brücke today equals that number of barges. In fact the numerical probabilities which Mills and Beatty recommend as fitness values are essentially meaningless except as they may be related to the numbers corresponding to other such probabilities. That relativity can be translated into a conjunction of Boolean statements without loss of empirical meaning so long as all possibilities are represented. For example, let us assume that the only litter size for a given species is one. We stipulate further that for this species, there is a thirty percent chance that an organism will not reproduce at all in its lifetime, a sixty percent chance that it will have just one litter in its lifetime, and a ten percent chance that it will have two litters. No other outcomes have ever been observed. Now if we have taken it as given that an individual having two litters is fitter than an organism which has just one, while this "one-timer" is fitter than one of its peers which fails to reproduce at all, then we have said all that we can say about the simple relationship among the three possibilities, and we have said it in Boolean terms.

### (3) Stand-alone versus grouped subjects of analysis

The third distinction is based upon what we can call "stand-alone subjects" and "grouped subjects," according to whether the units of analysis are taken as basic or not. Any given individual organism living in, say, the Jurassic period may have been fit or unfit compared to another individual in the same or a different species. The fossil record does not allow us to say much about the fecundity of individual organisms (apart from cases where it could be seen that an individual carried a certain number of eggs in its body or possessed apparently intact reproductive organs), even though we believe that individual organisms were indeed more or less fit as compared with one another then as they are now. The only level at which judgments of fitness can be made in this case is the group level, and thus the concept of fitness employed may be called "grouped fitness." Looking backward in time in other contexts, however, we can say that a given individual was or was not fit. For instance, the careful records maintained by animal breeders may indicate that some organisms were *in actuality* very fit, while others (sterile hybrids such as mules come immediately to mind) definitely were not. Here we can clearly talk about the fitness of the individual organisms described in the breeding records. In other words, a concept of "stand-alone fitness" has been employed.

At first glance it might be considered more appropriate to characterize this distinction with the terms "group fitness" versus "individual fitness" instead of using the unusual terms “grouped” and “stand-alone.” But this latter set of terms helps us avoid begging the question as to units of selection by assuming that the individual organism is that unit. Moreover, it should be clear that for the purposes of one researcher, a given unit of analysis -- the individual organism, say -- may be taken as the stand-alone level, while for another researcher only grouped fitness may apply to that level. In other words, one researcher might have information relevant to fitness at one level (again let us say the individual organism) and not be at all concerned about fitness as it might apply to lower, constituent levels (such as genes). In such a case, the sense of fitness is “stand-alone.” However, the next researcher might long to consider the fitness of the constituent parts without being able to do so. No doubt many paleontologists would be fascinated to know whether a one-of-a-kind fossil represents a particularly robust or a rather puny representative of its species. However, if the fossil record is sparse and the species seems to have gone extinct sometime between the days of the fossilized organism and the present, the best that can be said is that the group, the species, was apparently unfit.

It may be helpful to consider how the distinction between group versus stand-alone standards of measurement plays out. To do this, we would first need to assume a unit of selection. For the moment, let us say that the individual is that unit. Further, we will assume that natural selection is the motive force driving evolution. Now we can begin considering various possible units of analysis -- the individual, the gene, the group, for instance.

First possibility: the unit of selection is the individual, the unit of analysis is also the individual, natural selection is the mechanism of selection. In this case, we can speak of what the individual organism as the unit of selection actually does -- how long it lives or how many offspring it produces -- or what it tends to do. But our knowledge of its propensity to live a certain length of time or to produce a certain number of offspring is based on a comparison between some of its phenotypes and a body of observational and experimental results yielded by other individuals and perhaps even other species. These results relate those phenotypes to actual reproductive results and longevities. In short, talk of actualities as well as propensities is possible and meaningful, but the propensities are just an induction away from the actualities.

As a second possibility, let us make the same assumptions except to take the gene rather than the individual as the unit of analysis. The gene is, metaphorically speaking, "below the level" of the unit of selection. What this means in more concrete terms is that genes are constituent parts of individual organisms and not vice versa (but cf. Hampe and Morgan's evaluation of the consequences of Dawkins (1976, 1982)). Accordingly we cannot speak of any tendency of the gene toward a certain degree of fitness independent of the context of the body in which it is found. A gene for hooves might lend greater fitness to a terrestrial creature but not to a marine organism, but hooves which are too small, in terms of "footprint," would be disadvantageous if the weight of the organism is too great. In short, just as an individual does not possess any absolute fitness, but is rather fit only within the context of a given environment, so a gene *per se* is neither fit nor unfit. The gene is dependent on two contexts, one might say: not only is it a part of an overall environment but also of an "organismic environment."

We might make the discussion a bit more abstract at this point by considering what happens in general when the level of analysis is at a different level than the level of what is understood to be the unit of selection. In our present case, the unit of analysis is the gene while the unit of selection is assumed to be the individual (admittedly without justification and merely for the purposes of the present discussion), but we can conceive of several other combinations. The important aspect for the moment is simply the difference between the two levels. Now there are two possibilities. Either the unit of analysis is "below" (i.e., constituent of) the level of the unit of selection, or else vice versa. In the case where the unit of analysis is lower than the level of selection (as in our example of the gene as compared with the individual), it is very difficult to attribute a fitness propensity to the unit of analysis. This is because of the contingent nature of the lower level as compared to the higher. A given allele may have a propensity to fitness in one individual of a species and the opposite effect in a different individual of the same species who is a bit differently constructed, or whose environment changes a bit. It seems possible, for instance, that the ability to ovulate regularly improves the fitness of human females to a certain degree. However,

> In ancient times, when the food supply was scarce or fluctuated seasonally and when breast milk was a newborn's only food, a woman who became pregnant when she lacked an adequate store of body fat -- the most readily mobilized fuel in the body -- could have endangered both her own life and that of her developing fetus and newborn infant.
>
> Indeed, one can speculate that females who continued to ovulate in spite of being undernourished left no viable offspring or did not survive themselves; they therefore left no descendants. (Frisch 1988: 88)

### (4) Consequences of the three distinctions

Now let us put these three distinctions to work and see if we can answer the question which heads this subsection, namely, whether employing a concept of fitness other than the propensity interpretation is as great a mistake as Mills and Beatty imply, or whether an actuality interpretation may in fact be appropriate in some cases. One way to tackle the question is to consider possible scenarios based on the three distinctions just drawn. I believe it is clear without the need for a lengthy justification that *some* backward-looking analyses can adopt either an actuality or a propensity interpretation of fitness; in some such backward-looking cases, however, a propensity interpretation is simply superfluous. This is particularly true when the analyst is making Boolean as opposed to graduated judgments which bear on fitness (e.g., "yes, the species is extinct" or "no, the species is not extinct") and when the unit of life being analyzed is grouped rather than stand-alone for the purposes of the specific analysis undertaken. (Compare this with a case in which a backward-looking, Boolean analysis is applied to a stand-alone rather than a grouped unit. If we were to assert, "The Jurassic *individual* represented by this fossil is dead," we would have said nothing about the fitness of either the individual, the species, or any other grouping of like individuals.) The reason why a propensity interpretation is not necessary in such cases is that the very notion of tendency as related to the unit of analysis has to do with an average or a mean. The question is, what can be averaged and under what circumstances? The first necessity is that we have a spectrum of measurement which is not wholly discontinuous. If the questions we pose have to do only with extinction or non-extinction, we will not be interested in judgments that a species is "sort of extinct" or "rather not." There is a discontinuity on the spectrum between extinct and surviving which rules out any average between the two. This does not rule out the possibility of imposing a second scale, one which is not so completely discontinuous, on the judgment that a single taxon is extinct or not. We could count up the number of extinct species and describe it as a percentage of all species, for instance. But that would be a graduated rather than a Boolean analysis, and in any such case the focus of the "overlay analysis" -- meaning the counting up and other mathematical manipulation of the results of the "foundation analysis" -- differs from the original point of attack. An analysis can treat propensity as a *limit* to which some dynamic process tends, which is what happens in the case of forward-looking analyses which seek to know what would happen if *all* the stand-alone members relevant to an analysis were to be tracked and if their progress in some aspect such as reproduction were to be recorded in a graduated fashion.

Mills and Beatty end their paper by asserting the universality of their conclusion that the propensity interpretation is the proper perspective on fitness and is indeed the one adopted by evolutionary biologists, if only implicitly:

> We chose an example of microevolutionary change, since we wanted the least complicated instance possible in order to illuminate the form of explanations utilizing fitness ascriptions. We know of no reason to believe that a similar reconstruction could not be given for the case of macroevolutionary change (1977: 286; Sober 1984b: 55).

This statement should drive home the points just made. In fact there is very good reason to distinguish between micro- and macroevolutionary phenomena (to adopt Mills and Beatty's terminology). An instance of extinction would surely qualify as a macro-phenomenon, but in many instances it would be pointless to demand that the comparative analysis of the deceased species and its peers be done in propensity terms. In such cases natural selection is not arbitrating in a subtle struggle between phenotypes or behaviors which are barely distinguishable in their effects over the short term. Rather, the judgment has been rendered. The extinct species was less fit than those closely related species which survived, period.

There is another problem with Mills and Beatty's contention that the propensity interpretation avoids the connection with reality that causes the kinds of paradox which Mills and Beatty strive to avoid. We can set up the problem by looking carefully at what the authors believe the relationship between fitness and phenotypes to be. They assert that real phenotypic properties "constitute" an organism's fitness. "Thus, melanism is one of many physical properties which constitute the fitness or reproductive propensity of pepper moths in polluted areas (in the same sense that the ionic crystalline character of salt constitutes its propensity to dissolve in water" (1979: 271; Sober 1984b: 43). The authors do not emphasize the point; in fact, the reader senses that they overlook its importance. Shortly thereafter, Mills and Beatty reemphasize that fitness is itself considered to be a property. We should bear in mind that part of the article's agenda is to clarify how working evolutionary biologists employ the term. "Evolutionary biologists often speak of fitness as if it were a phenotypic trait -- i.e., a property of individuals" (1979: 272; Sober 1984b: 44). Mills and Beatty acknowledge this interpretation as legitimate.

Perhaps it would be well to restate the same points in a slightly different fashion. The combination of backward-looking, Boolean, and grouped analyses lends itself to an actuality interpretation. When these types of analysis coincide, a propensity interpretation is essentially superfluous. By contrast, the combination of forward-looking, graduated, and stand-alone analyses more or less demands a propensity interpretation. But rather than wading through a lengthy verbal description of all the possible combinations, perhaps it would be easier to understand the substance of this basic idea by looking at a graphic representation.

| | forward-looking | | backward-looking | | |
|---|---|---|---|---|---|
| stand-alone | PI* | PI | RI** | PI or RI | Boolean/graduated |
| grouped | PI | PI | RI | PI or RI | Boolean/graduated |

* PI ~ propensity interpretation　　**RI ~ reality interpretation

What this table shows is that whenever we speak of fitness as being revealed by events which have not yet occurred, or when we look at past events about which we have insufficient information to say with certainty what occurred afterward, then a propensity interpretation of fitness is appropriate. Any talk of future events or of events for which data is sketchy must be *conditional*; with respect to the future we can talk about a range of possible outcomes, but not of *the* outcome, and with respect to past events for which we have insufficient data we can speak of reconstructions which fit the facts at hand, and probably there will be more than one such. Thus forward-looking perspectives require us to talk of tendency rather than actuality, and so "PI" (for propensity interpretation) fills the entire column. On the other hand, there are times when we look at past events in such a way that no gaps remain: we know for a certainty what happened, but such certainty of course depends on the questions which we ask. If we ask whether some species is extinct, then it is possible that we know the answer to the question beyond a reasonable doubt. By contrast, we might ask how many individual members of the species existed at any given time, in which case we will again be uncertain. In such an instance we will simply have to guess by saying what the range of possible answers is, and perhaps we will have enough information as well to state a probable answer. But probabilities are based on numerical analyses of given data, analyses which amount to ratios or averages. When it comes to speaking of fitness as revealed by numerical average, we already know what perspective of fitness must be adopted -- a propensity interpretation.

It has already been indicated that sometimes the propensity interpretation does make sense, while understanding fitness only in terms of what really happens would be unwise and could lead to problems which would affect the whole of the theory of evolution. Under what circumstances would this be true? In accordance with our distinction between forward- and backward-looking analyses, it seems clear that a forward-looking purpose cannot take an actual past event as the definitive indicator of an organism's or taxon's fitness. In the sense used here, forward-looking does not imply that one analyzes the future from the vantage point of the present. That might be the case, but it is also possible to consider forward-looking as applying to possibilities. Take Mills and Beatty's example of identical twins, for instance, one of which is struck and killed by lightning. Even after the deadly bolt has done its damage, we can look forward in an imaginary way at what the deceased twin *might* have accomplished in the way of reproduction. Our primary tool in carrying out such speculation is of course comparison, and the units of comparison must be things like genotypes and phenotypes, and in short a “portrait” of the organism’s or taxon’s prospects based on observations of actual outcomes.

# Chapter Six: Can Fitness as Non-Circular Propensity Succeed?

As we have seen, the propensity interpretation of fitness proposed by Susan Mills and John Beatty (1979) offered a seductive way out of problems surrounding older definitions of the concept. These had appealed to differentials in real reproductive success or the supposed suitability of certain traits in a given environment. As Mills and Beatty rightly pointed out, such accounts seemed to falter in the face of counterexamples, such as those demonstrating that the intuitively more fit can enjoy less real reproductive success than the presumably less fit.

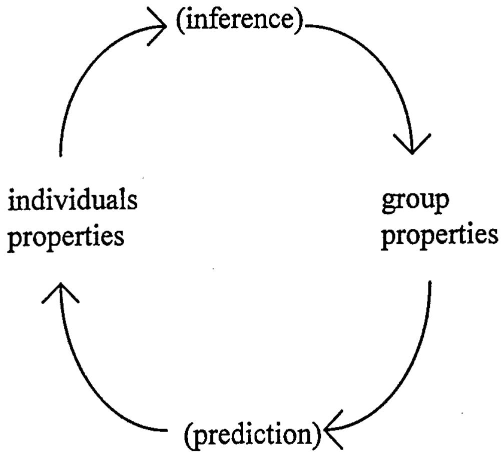
It remains problematic, however, whether the propensity interpretation escapes these same problems. A propensity can be thwarted (prevented from reaching its fruition) just as reproductive success viewed as a probability can fail to materialize. The concept of a propensity (which we will use interchangeably with disposition or tendency) would seem to find its foundation in empirical observation of numerous instances whose essential elements are the same. To make this case certainly requires a more thorough exposition, but perhaps it can be allowed without extensive argument that a propensity, though its existence may in some sense be independent of our knowledge of it, is discovered by us only after the fact--as what is *normal*. The normal, in turn, is established through recurring patterns constructed by observation of many individuals. As discussed above, the problem is then to decide what behaviors are relevant to the norm by which fitness is to be defined.

By this reasoning the propensity account does no more than establish a sort of "middleman" term which only defers identification of an empirical underpinning for judgments of fitness in the form of varying reproductive capacities or something similar. If this kind of empirical, measurable foundation leads to what Mills and Beatty call "inconsistencies" -- such as the intuitively more fit sometimes failing where the supposedly less fit succeed -- we may simply have to recognize the agency of chance (which arguably creates exceptions in all sciences) as something we cannot defend against. This chapter can therefore be read as a kind of lament: Would that the notion of propensity were clear-cut (but it's not)! Would that we could establish a clearcut, unique notion of probability which corresponds to an observed reality rather than one which is merely coherent with a certain perspective (but we can't)!

## 1. The units of propensity judgments

### (1) Individual versus "cumulative" propensities

A major problem confronting the propensity theory stems from the attempt to divorce *tendency* from what we can call "cumulative identities." In fact the cumulative sense of a norm is the only means of establishing the notion of a tendency which can be shown to exist in individuals. In so far as observers can know a certain tendency, it is a reversal of the real relation between group and individual tendency to claim (as Mills and Beatty do) that "a notion of fitness which refers to types...cannot be a propensity...[but] is a derivative of individual fitness propensities" (1979; Sober 1984c: 44). To restate yet again, from an epistemological standpoint, the cumulative tendencies of members of a taxon *can* be prior to our judgments of individual properties if we infer the latter from the former. A zoologist may look at a wolf pup, for instance, and predict that the animal will weigh nearly two-hundred pounds when fully grown. Such a prediction takes it that the properties of the individual will mimic an observed tendency conceived as the past performance of a taxon with respect to a particular characteristic. This is not to deny that an individual may have a propensity, unique to itself, of which we are unaware. Nor is it to deny that individual properties *can* be seen as prior to cumulative tendencies. The matter depends on which side of an essentially circular process we adopt as our observational standpoint.



To clarify further, a brief exposition of terms may be helpful. Our common sense can grasp three sorts of ends and corresponding conceptions of chance which we might tentatively describe as cyclic, unique, and extra-formal. The goal of drawing this distinction is not to offer a rigorous argument that one division rather than another is either sensible or exhaustive, though all the better if this classification proves to be either or both. Rather, the division will (with luck) yield a backdrop for a more rigorous discussion of fitness.

### (a) Cyclic ends

This first sort of end is a point in a pattern of similar motions such as the set of traits repeated across generations of a plant or animal kind. In such cycles not only the end but also the developmental process moving toward that end is the same or similar in one cycle as compared with another. An example of a cyclic end would be the adult form of a given species: among the individuals of each generation enough properties are the same to warrant our judging the adult as being or at least as indicating the "form" of that species for purposes of classification.

Admittedly this leaves underdeveloped some important issues -- what relation the adult individual bears to the form (identity or some more abstract relation), for instance. It is likewise unclear whether many developmental stages (insofar as each may be unique to the species) are as much indicative of the species' form as the adult is. But for present purposes it suffices to allow (a) that this type of end recurs; (b) that the end is both stopping point (*telos*) and formal cause in so far as one cycle is the immediate agent of the next cycle's creation (i.e., each cycle's existence is more than accidental with respect to the preceding and following cycles); (c) that the form of the cycle is *usually* attained ("always or for the most part," to coin an Aristotelian phrase); and (d) that we cannot resolve the question of whether an end or form of this kind exists independently of its instantiation. For simplicity's sake we are avoiding consideration of what might be called an epistemic, selective component in patterned motion. Put another way, we are not answering the question of whether a pattern can exist independently of someone perceiving it. Consider some relatively complex phenomenon such as a piece of music -- F. Scott Key's "Star Spangled Banner," say. We ordinarily hear arrangements of the piece which are of the John-Williams'-Boston-Pops variety. We recognize them easily and unanimously. But if we look hard enough -- in a retirement community, perhaps -- we will find someone who says that as a sequence of sounds, Jimi Hendrix's electric guitar arrangement bears no identity relation to the sequence wafting over Boston. The problem is one we have already visited above: how do we choose facets of a phenomenon to single out as essential and how do we describe those aspects? Simpler examples such as heavy bodies tending to be down seem less problematic, but it would still take a good deal of argument to rigorously resolve all ambiguities. If we chose to follow up on the possible epistemic facets of cycles or patterns, we might be led to a more idealist account than that just offered.)

### (b) Unique ends

We can also conceive of a second class of ends -- those recognized by some means other than their recurrence. An end of this type could be the unique (one-time) termination of a linear ascent -- of the universe as a whole, say, if one believes in what has been termed a "perfection principle." But commitment to such a strong thesis is not necessary. George Gaylord Simpson disparaged assumption of such a principle:

> "Examination of the actual record of life and of the evolutionary processes as these are now known raises such serious doubts regarding the oversimple and metaphysical concept of a pervasive perfection principle that we must reject it altogether....A description of what has occurred in the course of evolution will

> not in itself lead us to the identification of progress unless we decide beforehand that progress must be inherent in these changes" (1976: 240).

However, a unique end need not be the kind of perfection principle which Simpson rejects. It could be the final form of a now-extinct organism if we assume that the adaptations which occurred during the existence of the species were on balance positive. Progress to this final, optimal form could presumably encompass cycles or other subcontexts of directed motion, but the final form can be recognized as distinct from any recurring forms and therefore as unique.

### (c) Extra-formal (or super-formal) ends

A non-cyclic end might also be of the sort which can happen always or for the most part (and therefore is not necessarily unique) but need not (and so is distinguished from cyclic ends). Socrates, for instance, might represent a sort of human ideal which few men attain even though they are indeed men. In the sense of transcending a "subform" such as the one which defines a species, this type of end is extra- or super-formal. It is arguably associated with a form of a higher order than the one indicated by the end of a cyclic process. Further, a cyclic form would be a prerequisite for the existence of a "super-form" just as Socrates had to be a man in order to be a philosopher.

### (d) Kinds of chance corresponding to the types of ends

We can imagine two broad categories leading to corresponding conceptions of chance. The first is of randomness defined as an utter lack of directedness. In this guise chance would have nothing to do with any of the sorts of ends sketched above except in so far as it is the complement of directed (end-bound) action. The other approach to chance is to treat it as describing a disjunction between some part of a process and an end. (Apparently this is the sort of randomness which comes into play in the examples which propensity theorists offer to argue against the use of real measurable phenomena such as reproductive records.) In this second sense we can attempt to discriminate between types of chance corresponding to our provisional delineation of ends.

Cyclic chance is that kind of randomness which causes a progression to diverge from its expected cyclic pattern. The appearance of a two-headed calf is an instance of this kind of randomness. We expect the pregnant heifer to deliver what is more or less a carbon copy of itself at an earlier stage of its life, and when that does not happen, we may say that the fact of the matter has failed to conform to that which transpires always or for the most part. (More will be said about chance in cylic processes in subsection (3) below.)

Finding a kind of chance corresponding to a unique end is much more problematic. If the end toward which a process moves is *wholly* unique, how can an observer recognize a failure to meet that end caused by the agency of chance? Apparently there is no such means of recognition apart from failed expectation. For instance, certain religious sects have predicted the end of the world and have awaited that event in accordance with their particular beliefs. The leader of the Münster Anabaptists, Jan Matthys, predicted that the world would end on the 5th of April, 1534. His own death on that date spared him the awkward moment in which other prophets have seen the allegedly appointed hour come and go (Boyer 1992: 59). Millerites in the United States, for instance, predicted first that the world would end in 1843, later that the final day would be October 22, 1844 (ibid.: 81). In such instances of failed prediction, sect leaders must experience a kind of hermeneutic challenge as they address their perplexed congregations. What does one say in such a situation? "Sorry, folks, no end of the world today, but for what it's worth, we might get a little rain later on"? That will never do. Instead, the explanatory challenge is met by appealing to an unforseeable agency which postponed The End. Apparently the kind of chance which could deflect a progression of events away from a unique end would have a similar sort of character, only without the apocalyptic overtones.

A kind of chance conforming to super-formal ends is also difficult to imagine. Periodically philosophers of Socrates' stature appear, and when they do, the occurrence itself seems almost random. We can neither force nor predict the event; we also cannot wholly explain it after the fact. Even in a highly regimented utopia such as Plato's Republic or Aldous Huxley's Brave New World, there arise individuals who transcend birth and training or who fail to live up to expectations.

Perhaps there is a genus of super-formal ends in which this random flavor is not automatically present. Phylogeny in Haeckl's recapitulationism, in which ontogeny recapitultes phylogeny, seems to be such a perspective: the progession of generations move forward in an essentially cyclical way, but somehow the constituent cycles are minutely but sufficiently perturbed so that they "spiral" forward into new species. But as in the case of chance as an obstacle to unique ends, in such a genus of super-formal ends it is difficult to conceive of a way in which we could recognize the agency of chance, assuming it were operative.

## (2) A numerical example

Here's a simple algorithm that allows attrition and reproduction of organisms -- piglets, say. The "piglets" are pennies, shaken from a piggy bank:

Algorithm A

Rule 1: Pennies which land tails-up are chance casualties of the environment.
Rule 2: Twice the number of pennies that land heads-up are returned to the piggy bank to be used in the next round.
Conventions: A "round" begins with the shaking-out of pennies and ends with the return of pennies (if any) to the bank in accordance with Rules 2 and 3 above. P denotes the total number of pennies in the bank at any given time. T denotes the number of tails and H the number of heads which lie on the table. Alternatives other than heads or tails are not recognized.

Note that this "algorithm" may terminate itself after one round (if all pennies turn up tails) or after any round thereafter. Conceivably, too, the algorithm might never terminate. Interestingly, the probable populations before and after the first round are the same. This is because we expect half the pennies to turn up heads and half tails. (Recall that the algorithm requires us to put back into the bank twice the number of pennies which turned up heads.) If this is true of the first round, then why not of every round thereafter? Reasoning thus, we would expect the number of pennies before and after every round and before and after any combination of rounds to be the same. However, should the improbable happen at any point -- should more or fewer of the pennies turn up heads in the first round, for example -- then our prediction of the results of following rounds will be numerically different (though fomally the same). Say the bank contains 20 pennies, 9 of which come up heads. Then 18 pennies end the first round and begin the second. A difficult question arises. Which of the following options is correct, given what we mean when we talk about the probable outcome?

(1) The result of (i.e., number of pennies remaining after) the nth round = result of the first round.

(2) The result of the nth round = result of round (n-1).

Another way of posing the question is to ask what sort of transitivity relationship holds between rounds. Given the conditions stated and one way of reasoning, we would expect that at any "stage" in the experiment (i.e., at the end of any round) the number of pennies in the bank would be equal to the number first found there (at the beginning of round 1). This prediction, however, is what we might call an "outside" evaluation, that is, an estimate made in the absence of any information about what has happened in the previous stages. But "inside" information about the stage immediately previous to the one whose outcome we wish to predict is indeed relevant, as is information about any previous stages. Each round has the potential to be a perturbation, a result which can and should change our expectations of future results. (If, in our algorithm, twenty pennies are tossed but, contrary to outside prediction, only three heads turn up, we will expect fewer pennies in the bank following upcoming rounds.)

In this simple sense, popular expositions of how Darwinian evolution proceeds and how it is evaluated may require refinement. "Chance has no memory," notes Dennett (1994, 54). That may be true of chance itself, assuming there is such a thing. Some of our *observations* of phenomena involving chance, however, certainly do have "memories." These are the observations informed by "inside" information of the type which does not lead the observer into the gambler's fallacy (denial of the independence of individual events). In fact, predictive power in our algorithm depends on taking into account -- "remembering," one might say --what happens at each trial.

Here is one scenario:

| Round | Observation Point (Inside/Outside) | Prediction before the Round ($PR_x$) | Actual Result of the Round ($AR_x$) |
|---|---|---|---|
| One | Outside<br>(Only the algorithm and the initial number of pennies $P_0$ = 20 are known.) | $\{H_1 = 10,$<br>$T_1 = 10,$<br>$P_1 = 20,$<br>$P_x = 20\}$ | $\{H_1 = 9,$<br>$T_1 = 11,$<br>$P_2 = 18,$<br>$P_x = 18\}$ |
| Two | Inside<br>outside | $\{H_2 = 9,$<br>$T_2 = 9,$<br>$P_3 = 18,$<br>$P_x = 18\}$<br>$\{PR_1\}$ | $\{H_2 = 1,$<br>$T_2 = 17,$<br>$P_3 = 2,$<br>$P_x = 2\}$ |
| N | Inside<br>Outside | $\{H_n = H_{n-1},$<br>$T_n = T_{n-1},$<br>$P_{n+1} = 2H_n,$<br>$P_x = 2H_n\}$<br>$\{PR_1\}$ | ? |

Each individual coin toss is indeed independent of the ones which took place before; in that sense, chance as embodied in the coin tosses has no memory. However, the cumulative results -- $P_i$ in the notation above -- are dependent on previous trials and therefore do have a "memory" of a kind. (Again "memory" must be understood in such a way that the trials are independent at the "atomic" level, that is, so that the gambler's fallacy is avoided.) But what can we say to make this sense of "memory" more precise? First, we might question how it is that the results of individual coin tosses are perceived as random but independent, whereas our *prediction of* the overall result (the number of pennies put back into the piggy bank, that is) depends on our knowledge of what happened in the immediately preceding round. Thus we avoid the tricky proposition that the individual tosses are independent of one another whereas the cumulative result is not independent of the things of which it is the accumulation. The cumulative result *could* be said to be independent, too, in so far as it the function of independent events. The goal of making predictions, however, requires us to presume cumulative dependence. Similarly, when we attempt to explain past results, assuming a mutual "dependence" of any two adjacent rounds is helpful. ("Dependence" here does not mean that the result of one event changes the rules controlling the outcome of the next event. Rather, the dependence merely has to do with the fact that the outcome of one set of coin tosses becomes the initial condition of the next set.) In our model, macro-dependencies (the relationships between the results of adjacent rounds) have to do with prediction and explanation, not necessarily with the "real" circumstances resulting from mutually independent "micro-events" (individual coin tosses).

So far, so good. There is nothing earth-shaking in any of this. On the contrary, our intuitive grasp of the process seems to be satisfied. We expect that individuals, whether pennies or organisms, will be independent (within some parameters, of course); on the other hand, we expect that a system of autonomous individuals will be analyzable by relating "snapshots" of its history to one another.

However, there apparently are mistakes a researcher might make in observing and evaluating models of this kind:

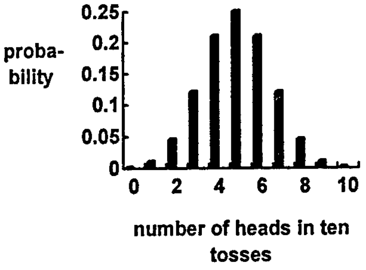
First mistake: Attributing the independence properly ascribable only to individual events to the system as a whole.

Second mistake: (the "vice versa" case) Attributing probability relationships properly ascribed only to cumulative states of the system to individuals.

From this perspective, let us consider teleological and causal explanations of biological phenomena. First, it seems clear that a system of dependent states, in so far as it is seen as such a system, is of course amenable to *causal* explanation. Here "causal" means simply that a state, a "snapshot," of the system may be described with reference to whatever has (a) occurred before and (b) could conceivably bear on the state in question. In our simple model, the overall result of Round 1 matters to the overall result of Round 2, at least in a probabilistic sense. In this case a "teleology" of the model, should anyone care to propose one, will always compete with a mechanistically causal account of how the stages relate to one another.

The level at which the teleological account does *not* have to compete with a causal one is that of autonomous individuals. Their independence as it relates to calculating probabilities is also a causal independence. It is not, however, a teleological independence. If one can discern a developmental pattern (a convergence toward an end point) among independent events, then presumably a teleological explanation would *in fact* be more potent than a causal one. But can there be such a convergence toward a goal among truly independent micro-events?

The graph below, following Kac (1964: 94), shows the behavior of coin tosses:



The probability of exactly five heads in ten tosses is 252/1024. (Here an event takes account of the order in which heads and tails appear as well as the respective numbers of heads and tails which turn up. So, for example, flipping four heads followed by six tails and flipping six tails followed by four heads are both ways of achieving exactly four heads.) Kac notes (p. 95):

> If we plot the same kind of graph for 10,000 tosses, it becomes much wider and lower: the high point (for 5000 heads) is not in the neighborhood of 25 percent but only 1/[100√π], or approximately .56 percent. (It may seem odd that in increasing the number of tosses we greatly reduce the chances of heads coming up exactly half the time, but the oddity disappears as soon as one realizes that a strict 50-50 division

> between heads and tails is still only one of the possible outcomes, and with each toss we have increased the total number of possible results.)

By the reasoning of the propensity theory of fitness, the identity of the "normal" surviving organism (within a given generation) is perceived not so much as an object of study in itself; rather the concept of the normal fit individual is known through convergence, through a process of calculation. Interestingly, it is David Lachterman's general thesis that the mathematics of modernity, perhaps modernity itself, deals primarily with process (construction, *poiesis*, *techne*) whereas classical mathematics studied nature itself (*phusis*) as an assembly of existing things. Probability as a branch of mathematics (which, paradoxically, Lachterman seems not to emphasize in his book) is of course very much a child of the Renaissance and Enlightenment, and true to Lachterman's generalization, one might rightly say that its methodologies emphasize process over object. One can see this in Kac's example, in which coin flips clearly *converge* (as process) toward a 50-50 split while that precise number (object) itself becomes increasingly less likely to occur. The "object" is known not by direct observation but instead by the *movements* of other objects which are themselves ephemeral in so far as they are taken as aspects of movement for the purpose at hand. One of course thinks of the calculus, the "theory of fluxions," in this regard as well.

### (3) Chance in cyclic processes

We recognize aberrations such as the ancient examples of the two-headed calf or man-faced oxen as the product of chance on the basis of two observations: (a) the source process is familiar to us (sufficiently similar in essence to past processes that we can say they are the *same* as the present process in a formal sense even though each concrete instantiation of that formalism may be viewed as an independent thing), and (b) a point in the process (the anomaly) differs from the analogous point in past occurrences of the formally identical process. But these classic instances play a role in cyclic processes only. It would be worthwhile for us to inquire about the relationship of processes to various ends in general. Let us call a thing similar or identical to its analogs (e.g., a normal calf in certain cyclic processes), an expected but unique thing (the end of the world in a particular, apocalyptically unique process), or a partly unique, partly expected thing (figures such as Socrates among generations of humankind) respectively a cyclic end, a unique end, and a super-formal end. Further let us accept that context determines what category of form a thing belongs to. If scientifically advanced extra-terrestrials were to snatch a pregnant woman for scientific examination upon their first visit to earth -- as happens rather frequently according to the kinds of newspapers seen around supermarket checkout lines -- they may not know what her baby will look like based on past experience (i.e., as a stage in a cyclic *process*), but they may be able to predict its nature based on their analysis of the fetus's biochemistry. In that case, they view the baby as the result of a unique *process*. One does not need to believe such stories (I certainly do not) to understand the basic point: the nature of a process (cyclic, unique, or super-formal) depends on one's perspective.

A brief summary of the relationships between various types of ends (or processes) and kinds of forms might go like this:

| Type of form<br>Type of process | Subform (SbF) | Unique form (UF) | Superform (SpF) |
|---|---|---|---|
| Cyclic (C) | (1)<br>$C \Leftrightarrow SbF$ | (2)<br>ambiguous | (3)<br>$C \Leftarrow SpF$ |
| Unique (U) | (4)<br>ambiguous | (5)<br>$U \Leftrightarrow UF$ | (6)<br>ambiguous |
| Super-formal (SF) | (7)<br>$SF \Rightarrow SbF$ | (8)<br>ambiguous | (9)<br>$SF \Leftrightarrow SpF$ |
| Propensity (P) | $P \Leftrightarrow (SbF \vee U \vee SpF)$ | | |

The existence of unique and extra-formal ends allows us to make one concession to the propensity interpretation. Presumably such ends would be independent of any manifestation or embodiment. (If not in a full-blown Platonic sense of independent forms, this type of end is independent at least insofar as it need not be materially realized at any given time.) In other words, a unique or super-formal end can be conjectured with no possibility of being tested. In particular, it can be conjectured of an individual without reference to any group in which the individual might find itself. It would then be proper to speak of an individual's propensity to a certain state of being independently of any corresponding (cyclic) propensity of the individual's type.

However, this sort of type-independent propensity is ill-suited for analysis by evolutionary biologists. Consider the case of two individuals, I1 and I2. Assuming that I1 has a certain propensity which makes it more "fit" (leaving the meaning of that term undecided for the moment) than I2, how will we recognize the fitness-giving propensity except in terms of an operative special quality, an attribute? We can choose any number of terms or euphemisms to describe what happens to the individuals as wholes, but we will find little gratification in treating organisms as "black boxes" whose inner workings are veiled. Instead, we will want to know what it is about certain organisms which makes them perform so differently than their close counterparts in the same environment that we are moved to make judgments of relative fitness. This knowledge requires analysis of the attributes of individuals, ultimately culminating in statements like: "If an organism has attribute a1 in environment e1, it is statistically more likely to reproduce (other factors remaining equal) than an organism which is nearly identical but does not possess attribute a1." In the end our judgment must be based on the test of a hypothesis, as in any scientific endeavor.

This last statement assumes a great deal about the nature of scientific inquiry without any supporting arguments. Perhaps this can be forgiven if we focus on the epistemic issue of why and how we attribute fitness. Studies of evolution admittedly are often explanatory rather than predictive, but this does not mean that their goal is only to make assertions about wholes with no regard to component attributes. We may say that the long-term survival of a given organism is evidence of fitness, but surely we will also speculate as to the particular traits of the species which allowed it to thrive in its specific environment. To introduce consideration of attributes in this way is to require tests even if of a different sort than laboratory experiments; tests, in turn, require repetition or type-based studies. (Type as opposed to individual, that is.)

This is not to claim that those evolutionary biologists who ascribe to non-propensity accounts of fitness either have or require foolproof methods of identifying the individual attributes on which the propensity to fitness of wholes depends. There are obvious actuarial instances of the problems inherent in separating aspects of an individual for the purpose of obtaining a quantified probability. Chance can distort the most careful quantification -- as in the case of the differently-marked butterflies mentioned by Mills and Beatty (1979: 273, n. 2; Sober 1984c: 40 - 41). The problem certainly is not limited to quantities nor does it apply only to easily-observed behaviors, as Suppes notes in his discussion of "invariance" and "segmentation."

> In principle, phonemic sound should be able to reduce the continuous sound wave to a finite linear sequence of phonemes....But problems similar to those besetting the claims for distinctive features are now well recognized. They are the problems of invariance and segmentation. Spectographic and other physical analysis of speech sounds shows that there is no straightforward identification of the invariants for a given short time interval that correspond to a given phoneme, and similarly there is no obvious, or even subtle, marking of the segmentation into distinct phonemic units. (1984: 137)

Roughly, for the purposes of evolutionists, the problem of invariance would lie in finding sufficient similarity among attributes to allow identification of the determinants of fitness. It is closely related to the notion of segmentation by which the "building blocks" of larger wholes -- the attributes of organisms -- are delimited. It would be correct to say that invariance and segmentation pose problems for traditional concepts of fitness based on traits or reproductive differentials. But that is not at issue here. What we need to investigate is whether the propensity interpretation is superior to its predecessors by virtue of escaping these problems.

### (4) Vague reference

Judgments of probability raise additional questions about the adequacy of the propensity account. Such judgments presumably depend on real-world observations and can lead to what Hempel calls the "statistical syllogism" based on the probability r of set G with respect to set F, denoted p(G, F) = r. (Since Mills and Beatty claim to establish a "Hempelian reconstruction" of fitness-based explanations of evolutionary theory (Hempel 1960: 462), it seems appropriate to appeal to another aspect of Hempel's work in this context.) The statistical syllogism takes the form:

> a is F
> The statistical probability for an F to be a G is nearly 1
> So, it is almost certain that a is G (ibid.)

Hempel points out that the class of syllogisms representable in this way can lead to "inconsistencies" because "the individual case a...will in fact belong to different sets." So, for instance, the probability that Jones is a millionaire depends on whether we consider him as a Texas oilman or a philosopher, even though he may be both. Hempel attributes the example to Barker:

> Suppose that Jones is a Texan, and that 99 per cent of Texans are millionaires; but that Jones is also a philosopher, and that only 1 per cent of these are millionaires. Then rule (2.1) ["a is F; The proportion of F's that G is Q; Hence, with probability q, a is G"] permits the construction of two statistical syllogisms, both with true premises, which yield the incompatible conclusions that, with probability .99, Jones is a millionaire, and that with probability, .01, Jones is a millionaire.

We can get widely differing probabilities depending on the way we identify Jones.

Ettinger et al (1990) offer a similar example[25] as part of their argument that fitness is properly ascribable to types, not to individuals. The premise of their argument is that we infer fitness by observing numerous individuals in a given taxon. While this supposition seems to be correct, however, it does not necessarily warrant the authors' conclusion. Depending on the definition of fitness employed, it may be that a given individual possesses a certain level of fitness but that, because of the difficulties in categorizing which Hempel and Ettinger et al have raised, we cannot know with certainty what that fitness level is. Or we could express the same thought like this: we cannot know an individual's fitness because an individual belongs to too many different categories (groupings which we can associate with various fitness levels); but the epistemological impossibility of reckoning and expressing individual fitness does not preclude the possibility that an individual *can* have a fitness value in a metaphysical sense (apparently contrary to the conclusion of Ettinger et al). The key issue is again not the level at which fitness can be ascribed but rather its basic meaning. Ettinger *et al* offer this definition:

> Type X is fitter than Type Y in environment e if and only if the organisms representing X are on the average reproductively more successful in environment e than those representing Y and if, furthermore, we can infer a causal relationship between the relevant hereditary physical properties of the individuals representing the types and the differences in reproductive success. (1990: 507)

In itself there is nothing wrong with this understanding of fitness. It is a way of tackling what we could dub the "level of fitness" problem which might be useful in some contexts. But contrary to the authors' claims, there is no need to exclude all possibility of talking about the fitness of individuals, nor is it likely that biologists and philosophers will give up that convenience. Rather, why not simply admit that it is difficult to infer the properties of categories from their members and vice versa and then seek a means of accommodating the difficulty? To find such a means we must recognize that the challenge is not to find which level -- individual or group -- is somehow privileged with respect to the task of reckoning fitness. What we need is a way of showing, in our definition of fitness, that individual and group bear some close and necessary relationship to one another, a relationship which allows to make inferences about fitness in *either* direction. Moreover, the definition of fitness should be abstract enough to accommodate talk of fit groups, individuals, phenotypes, genotypes, or whatever other units might serve the purposes of research and dialogue. It seems to me clear that a recursive definition of the form $\text{fitness}_t = f(\text{fitness}_{t-1})$ fills the bill. By emphasizing the aggregate character of fitness, such a definition permits the analyst to trace a relationship between individual data relevant to fitness and higher-level "fitnesses."

Two brief reflections may help to clarify my point here. In the piggy bank scenario above (pp. 189 ff.), tails are wholly unfit while heads are as fit as an individual or category can be. (Recall that when a flipped penny turns up heads, it "reproduces" in the sense that it causes another penny to be put into the population.) In this example there is clearly no problem in categorizing individuals: a penny which turns up heads belongs only to the category of heads-up pennies. Thus in such a "pure" scenario, it is no surprise if we can reckon the fitness of the aggregate case by looking at what happened to the individuals constituting the collective under consideration. But what happens when an individual belongs to more than one category (the situation which concerned Hempel and, for a somewhat different reason, Ettinger *et al*)? The answer to that question depends on how we have chosen to define fitness. The type-casting problems which Hempel and Ettinger *et al* point out can be wielded against the claim that the propensity theory of fitness makes no use of the kind of cumulative empirical content which is available through a conception of cyclic (rather than unique or super-formal) ends. To see this, we should first acknowledge that we can treat both an organism and its environment -- in so far as we know them -- as *conjunctions* of attributes. Some of these attributes are essential: for instance, mammals must be hair-covered (a "categorical" or "definitional" requirement) and they must be vertebrates (an "inclusive" requirement, expressing a necessary relation between sets) (Klenk 1983: 211).

However, this "foundation" of essential attributes is probably of only peripheral interest to the many discussions of fitness and selection which are not interested specifically in taxonomic divisions. It is not clear whether the essential attributes of the class Mammalia have an analog in the other characteristics of an individual mammal, and in fact even the criteria for belonging to the class of mammals allow some leeway, as a list of the defining characteristics shows: "Seven cervical (neck) vertebrae in *most* species...Two pairs of limbs for locomotion in *most* species" (Otto and Towle 1973: 596; my emphasis). In questions of evolution, most of the attributes which interest us will bear no necessary relation to the individual organism and therefore can be selected in much the way as "Jones" in the example above can be considered as either a Texan or a philosopher. In other words, we are able to refer to the organism adequately without simultaneously listing the traits which particularly interest us as signaling a change in fitness compared with that of its ancestors.

The challenge then becomes to assign probabilities (which amount to a means of expressing the differentials encapsulated in terms such as "survival potential" or "reproductive potential") to conjunctions of attributes. But how can this be done without recourse to observation of the empirical evidence in the form of individuals possessing all or subsets of the attributes under consideration? If there were some other means, evolutionary biology could transform itself into a powerfully predictive science -- a status few if any practitioners would claim for their subject. The possibility of assigning such probabilities at all is a matter for debate, but even if statistics can in fact be gathered, their source must be a matter of observing cumulative rather than individual propensities.

Even those who claim we can "assemble" the probability of a conjunction from the probabilities of its constituents hesitate to allow the "conjunctive" probability independence from evidence. Berenson raises an interesting example based on the question of whether we can combine actuarial statistics in such a way that they are relevant to conjunctive cases which we have not yet encountered or at least have not studied scientifically for the purpose of evaluating probability. What statistical relation, he asks, does the mortality of smokers and the mortality of drinkers bear to the mortality of smokers-who-drink if we have empirical evidence for the former two cases but not their conjunction? (Berenson 1984, 28-34). Following the same line of questioning, the evolutionary biologist can ask whether any example of fitness and adaptation is comparable to any other when the organisms involved are treated as wholes. This daunting variability of organisms and their relationships to selective environments (involving "macroscopic versions" of the problems of invariance and segmentation) led Mills and Beatty to reject a trait-bound theory of fitness (Mills and Beatty 1979; Sober 1984c: 41). But the same problems afflict the study of individuals with respect to their traits.

## 2. A timeless smile (logical relation versus finite frequency accounts of probability)

We have seen that a questionable aspect of the propensity theory of fitness proposed by Mills and Beatty is its wavering attention to chance as a disrupter of definitions of fitness. On the one hand, these researchers recognize that the possibility of aberrations defeats definitions of fitness previously held. They cite (Sober 1984c: 40) Scriven's case of the "lucky twin" to show how what we think should be a link between individual attributes and reproductive success can fail -- so that the twins' prima facie equality crumbles under certain "chance" conditions. Similarly, they offer an even stronger example where a butterfly which we "sense" is less fit (based on what we know of its attributes and of its environment) than another butterfly nevertheless has greater reproductive success, again by virtue of chance. In view of such examples and the accompanying analysis, it would seem reasonable for Mills and Beatty to ground their "propensity interpretation of fitness" in such a way that norms constitute a sort of buffer against chance. It is the abnormal case, after all, which the propensity theorists see as invalidating the earlier definitions of fitness in terms of traits empirically associated with reproductive success in the past.

But instead the propensity theory seems determined to avoid any account of chance at all. It is as though the propensity theorists feel there is a continuity in the relation between the individual and its traits which does not apply to populations of individuals in such a way that reproductive success (e.g.) is a viable criterion for fitness: "propensities are dispositions of individual objects.... Classes -- abstract objects, in general--do not have dispositions, tendencies, or propensities in any orthodox sense of the term" (Sober 1984c: 44). As we have already suggested, this seems to be precisely backward. It seems more correct to say that classes are the primary possessors of tendency in so far as dispositions are defined only in terms of the sort of cyclic generality which cuts across instants of time and individuals of a kind. In fact we must look beyond the individual to establish what its tendency is, and we should bear in mind that a "tendency" is yielded only by abstracting certain attributes of individuals from observations of patterns, i.e., cycles. Mills and Beatty reject trait-bound accounts of fitness partly because, as they say, "The features of organisms which contribute to their survival and reproductive success are endlessly varied and context dependent" (Mills and Beatty, 41 in Sober 1984). What precludes our substituting the term "propensity" for "features" in this statement? Apparently nothing.

Let me approach this in a slightly different fashion. If we choose to understand judgments of probability as being finished works, as a painter may call a certain picture finished and then move on to the next project, then apparently there is no way that a judgment of propensity can fail. It has been argued that probability judgments in general are based on evaluation of a finite set of circumstances. This view has been associated with Reichenbach ($^{2}$1971) and may be attributable to a general logical empiricist goal of translating all non-analytic statements into propositions relating finite collections of empirical facts. Berenson cites Reichenbach's assertion that "[i]t is with the sequences having a practical limit that all actual statistics are concerned" (ibid.: 348), then counters:

> This proposal obviously turns on Reichenbach's claim that finite sequences 'exhaust all the possible observations of a human lifetime or the lifetime of the human race', and it is clear that Reichenbach has in mind here what I have called limited finite sequences; his claim is that experiments exceeding in number some fixed amount would be 'inaccessible to human experience.' Indeed he cites the figure of billions of elements as being beyond the domain 'accessible to human observation'. However, the main thing which limits the domain accessible to human observation is time, and Reichenbach's argument would only succeed if there were a fixed finite limit to the lifetime of the whole human race, which Reichenbach appears to believe. (Berenson 1984: 197).

While it is not so clear as Berenson would have it that time (rather than, e.g., finite human cognitive ability) is the primary limiting factor in this context, his general critique of Reichenbach makes a good case for the work-in-progress character of probability judgments.

Indeed, it seems desirable to treat probability judgments as transcending any finite collection of specific facts. Some works of art are never finished -- "Like Kafka, whom he resembles in many ways, Leonardo certainly found it hard to finish his works: he did not complete the Sforza monument, the *Musician* in the Ambrosiana, the *Battle of the Anghiari*, the *Virgin with Saint Anne*, *The Last Supper*, or the *Mona Lisa*" (Bramly 1991: 166) -- and likewise probability judgments may always be subject to revision and reinterpretation. Opposed to a finite frequency theory as developed (e.g.) by Reichenbach is the *logical relation theory* (l.r.t.) of probability, which Berenson describes as

> the view that probability judgments are always relative to evidence, and in fact merely indicate the degree of certainty a particular body of evidence gives to the hypothesis we are concerned with....
>
> The most prominent feature of the l.r.t. is that, on it, no hypothesis can be regarded as probable in its own right; rather, it is a hypothesis coupled with some body of evidence that alone can be said to be 'probable'. (ibid.: 13 - 14)

The l.r.t., it should be noted, is not necessarily hostile to all brands of logical empiricism simply because Berenson opposes it to Reichenbach's finite frequency theory. That is obvious in so far as Berenson classifies Carnap's theory of probability (as well as Keynes') as a "version" of the l.r.t.

This is not the place for a lengthy exposition of the l.r.t., but it should be clear how well Berenson's brief description of it (quoted above) jibes with the notion that a recursive interpretation of fitness (or indeed of any quality in which chance plays a role) must be considered (1) as a work in progress and (2) as a statement about quiddities in which none of the three elements -- neither argument, function, nor value

-- can be considered in the absence of the other elements. The reason why a propensity à la Mills and Beatty's development can fail is that it violates (1) and (2) above. Mills and Beatty seem to treat propensity-based fitnesses *qua* internal properties of an organism as finite, finished entities. Moreover, they do not distinguish between the probability judgment as a formal algorithm (the *function*, in Frege's terminology) and its manifestation at any point in time based on data considered up to that moment (Frege's *value* of the function).

## 3. Two possible bases for defining fitness: traits and reproductive success

Mills and Beatty reject the notion of a trait-based definition of fitness (1979: 41 in Sober 1984c). Their reason for doing so, however, seems misguided. Moreover, it may be the case that biologists in fact do adopt a trait-based definition of fitness, at least to some extent. Let us take these matters one at a time.

First, Mills and Beatty appear to miss the point of a trait-based definition of fitness when they assert:

> ...[N]o one has seriously proposed such a definition, and it is easy to see why. The features of organisms which contribute to their survival and reproductive success are endlessly varied and context-dependent. What do the fittest germ, the fittest geranium, and the fittest chimpanzee have in common? (Sober 1984c: 41)

Contrary to Mills and Beatty's implication here, the point of a trait-based definition of fitness is not to posit a unifying "fitness factor" common to all organisms in all environments. Rather, the goal of basing the definition of fitness at least partly in traits rather than wholly in reproductive success would be to avoid the sort of counter-intuitive conclusions which Mills and Beatty themselves recognize (Sober 1984c: 40-41, esp. footnote 2), for example in instances involving the chance destruction of a pre-reproductive individual which intuitively is as fit or even fitter than its contemporaries. A definition of fitness based on reproductive success forces one to conclude, counter-intuitively, that the individual which dies before reproducing is in fact less fit than its contemporaries.

Second, Mills and Beatty seem to disregard the possibility that biologists in fact do adopt a trait-based notion of fitness. If this were not true, how is it that a researcher would recognize the agency of chance in explaining the differential reproductive success of twins in a scenario such as the one which Mills and Beatty adopt? Recall one of the primary examples:

> Scrivens (1959) invites us to imagine a case in which two identical twins are standing in the forest. As it happens, one of them is struck by lightning, and the other is spared. The latter goes on to reproduce while the former leaves no offspring. (Sober 1984c: 41)

The notion of a tendency depends on identification of that to which a certain process tends. In other words, the concept is inevitably teleological, if only cyclically so. If we define fitness in terms of an individual's tendency to reproduce successfully, all the same questions which confronted earlier definitions of fitness in terms of species-based reproductive success remain, though perhaps once removed. The criterion of fitness must be recognized first; then, by a process of backward extrapolation, the tendency is sketched out as a sort of continuum between the organism and its standard of success. Whether fitness is seen as exclusively a nebulous potential of the individual or is allowed expression in type-based criteria such as reproductive success of a species, the process is liable to various pitfalls (particularly the agency of chance).

A propensity theorist might make an argument for what we can call continuity in change -- whether that means the growth of an individual or the evolution of a natural kind. The motivation would be to show how, based on an initial random mutation, the individual acquires a property which can be perpetuated but which can also be studied apart from its recurrence in succeeding generations, and which is not vulnerable to the operation of chance which afflicted the twins and butterflies mentioned above. If we consider the ascent to realization of the adult form of an organism as divisible into discriminable "snapshots," then the state immediately preceding complete fulfillment of that form is arguably very close to being *necessary*. In other words, the reproductive potential of the individual is very nearly realized.

(It may seem ill-advised to employ a spatial metaphor in questions of necessity and accident, but perhaps we can allow such a discussion as a starting point. It may even be true that the process of maturation or of evolution is *sui generis* and defies description except through metaphors, in which case this terminology is the best we can do.)

We should bear in mind that if the progress toward "form" (a shorthand we can use for "adulthood") is continuous in the sense of contemporary mathematics or of Aristotle's *Physics* V. (227a10ff), then we can pick a state "infinitely close" in time to the realization of form. Now although it may not be necessary and sufficient to the achievement of the form, that next-to-last state is at least linked to the form *qua telos* by some real facets of the organism and its relation to its environment.

For argument's sake, suppose we could make the same inference we have just made with respect to the next-to-next-to-last state, then with the state preceding that, and so on. In other words, we would extrapolate along the continuum away from the form in infinitely small increments. Further, we would claim that this regression is not a matter of interpolation between the end and the beginning of the process since our aim is to claim a near-necessary relationship between states on the basis of extreme nearness. The alternative would be to argue on grounds of position between two distant points (e.g., child and man) which are thought to be linked.

Does chance of the kind which Mills and Beatty use to preclude acceptance of classical views of fitness force a break in such a hypothetical chain? That is, can we extrapolate all the way down to the "beginning" state (the infant or even the fetus, say) or must we stop at the point where the instance of chance "bends" the causal sequence toward an unanticipated end which has a different "fitness value" than the one expected? It seems the only possible answer for the propensity theorist is that at no point will our toilsome regression encounter a gap -- a sort of vacuum where maturation stops or is at least suspended -- and so we can indeed proceed all the way to the beginning of the process. To posit a conceptual precipice analogous to a pre-Columbian mariner's "edge of the world" past which one cannot sail is to invite the challenge of explaining how such gaps are bridged (as they surely would have to be if we wish to maintain that growth is continuous).

Now there are two complications for the propensity theorist here. First, it appears that the same argument as above could be made for units of study other than the individual. For them, as for the single organism, the only way to eliminate the possibility of making such a backward progression is to deny a necessary relation between states of growth or evolution no matter how close they are. But if there is no essential cohesion between any two states (or points in our spatial metaphor) regardless of their nearness, then from a "microscopic" perspective chance differs from regularity only as it does in a "macroscopic" scenario: in cyclic processes *frequency* is antithetical to chance, while in unique chains we simply take our best guess as to "what should have happened" given initial environmental conditions and our understanding of the significance of individual traits.

The second complication is that our assumption of continuity introduces an element of indeterminacy, no matter if the change considered takes place in the individual or in some wider context. To see this, let e be a state which can result in one of a number of states in the set F: {$f_1$, $f_2$, $f_3$...}. (We defer consideration of whether F is finite or infinite.) In turn, each $f_i$ in F can be followed by a set of consequences: f1 by G: {g1, g2, g3,...}, f2 by H: {h1, h2, h3,...} and so on. (There may be some overlap between sets.) We can call the set of such sets (i.e., the set of consequences of F) S: {G, H,...}. In short, then, we have ei---->F---->S, where S is a sort of "super set," or in other words, a set of sets.

Now there is a controversial assumption underlying this reasoning. Roughly, it is that even in the absence of a deterministic link between states, complexity increases in a rather orderly way: we link each state with a set of states rather than with a set of sets of states, or of sets of sets of states, etc. If that is consciously a matter of convention alone, there is no harm done. But if one holds this as a necessary way of viewing growth or evolution, while simultaneously opposing determinism and championing continuity, an inconsistency emerges. To see this, assume $e_i$---->F as above. Assuming continuity with no determinism, there is a set F' intervening between $e_i$ and F. (If there were nothing intervening, we would lose continuity; if that which intervenes is ever a single state rather than a set, we have a deterministic relation between $e_i$ and that state.) The imposition of a set of possibilities between "points" of change reintroduces Hempel's problem: Which of the elements of the intervening set is the proper "defining" quality of the organism?

So far this discussion has treated a consequent as a disjunction (e.g., $e_i$ leads to $f_1$ or $f_2$ or $f_3$ or...), but that is not to consciously ignore our inclination to treat some states, particularly ends, as conjunctive sets. We might define a human being, for instance, as a thinking animal, i.e., as {a thing that thinks, [and] a thing that is an animal}. This should not cloud the issue we are discussing, since even if a "single state" is taken to be a conjunctive set, it is still distinguishable from the sort of disjunctive set we have established as consequents.

Relative proximity of states is apparently not the way to make rigorous the distinction between significant and insignificant attributes, even though we may feel the link between certain traits and some measure of fitness (e.g., survival potential) should be weaker than that joining another aspect with its possessor. (The organism might have the same life expectancy in a given environment regardless of whether it is a bird or a mammal). We may feel certain that one state is integral to fitness while the other is merely parasitic on the benefits reaped from the presence of its "colleague," yet there is a frustrating continuity between them. A similar argument could be made with respect to the other side of the causal chain -- that is, to a kind of formal cause (to use Aristotelian language) as a motive principle, an impeller from the first state in a sequence of growth until the last, when form is realized.

Put more simply, one feels there is a distance amounting almost to a disjunction between a goal and a chance mutation which happens to contribute to its achievement, yet the beginning event in a chain diverted by chance is also in some sense remote from the chain's ultimate end. Neither the end nor the beginning could be called random, though their relationship might fit that description. Two further observations inexorably follow: first, there cannot be a complete discontinuity between the instance of chance and the end which it happens to have a part in realizing (else we would not say that the chance event was indeed part of the causal chain leading up to that end); and secondly, the difference between the chance event and others in that chain is not one we can specify in a rigorous, qualitative fashion solely with respect to the *telos* in question. Individual and cumulative changes (those expressed in what Mills and Beatty call types) are exactly the same in this respect.

The weakness of the extrapolation approach points out that continuity is a two-edged sword. On the one hand the concept of a continuous "line" of change gives us confidence that there is no gap intervening between two points or states under analysis. On the other hand the attempt to establish a necessary progression of states on the basis of nearness fails because we can conceptually abut two states only provisionally: continuity insists that there are additional points (states) between any two no matter how near the elements of that pair. The lack of such a necessary progression leads to a situation where we cannot discern whether an organism is or is not more fit than another of its kind without appeal to more long-term observations of types.

Even if we decide to ignore intervening states and allow that two states are "adjacent" by convention, it is not clear that there is any non-conventional necessity between them. Perhaps the best we can do is to allow for relative frequency, which is established by observation of types rather than individuals. The question still remains whether we prefer to claim that conditional statements are never absolutely true (but instead indicate at best relative frequency) or that there may be a class of events which are necessarily tied together. To explain the occurrence of the irregular changes (mutations) which lead to improved fitness we appeal to coincidence, to chance. Our evaluation of fitness potential with respect to the attribute is never conclusive, since there is always the possibility that a wider context will disprove our predictions or that a pattern of improved fitness will emerge from data where it had once seemed no such pattern could be found.

## 4. What kind of force is natural selection and how does it relate to fitness?

The propensity interpretation of fitness is valuable in underscoring a "mystery" associated with the term -- how organisms which we call "fit" can sometimes fail to thrive as expected, whether through some unseen internal defect or some unanticipated agency of environment. However, exceptions are no reason to abolish rules. Almost any attribute allows aberrations: "Smart people sometimes say dumb things," as the saying goes, but that is no reason to toss out all our past conclusions about what "smart" means. The propensity theory shares too much with the accounts it intends to supplant to be a viable replacement. In particular, its meaning is still based on the evaluation of patterns of behavior and it is still subject to exceptions stemming from chance occurrences.

Perhaps the best way of putting all foregoing objections to the propensity interpretation into a nutshell is to reflect that even the broadest-brush theories of evolutionary trends -- theories which are clearly based on alleged tendencies rather than observed realities -- can be contradicted. One hears, for instance, that the tendency of evolution is toward complexity. Perhaps, but it is also possible to consider that evolution is directionless except in the sense of seeking *increase*. Gould considers the theory that the spectrum of life on earth begins at a "left wall" separating the most simple life forms from non-living matter and extends (in swiftly diminishing numbers) toward the right in the form of increasingly complex organisms. With qualifications, he rejects the assertion that the "direction" of evolution favors complexity:

> "...When we consider that for each mode of life involving greater complexity, there probably exists an equally advantageous style based on greater simplicity of form..., then preferential evolution toward complexity seems unlikely a priori. Our impression that life evolves toward greater complexity is probably only a bias inspired by parochial focus on ourselves, and consequent overattention to complexifying creatures, while we ignore just as many lineages adapting equally well by becoming simpler in form" (1994, 87).

Similarly, Gould notes that mammals currently prevail while dinosaurs are extinct not because the mammals were in any sense superior to them during their period of coexistence, but rather due to the "fortuity" of some irregular event (probably the impact of a huge asteroid 65 million years ago; ibid.). Rather than taking the intrusion of chance in this instance as another argument for a propensity interpretation, we should ask whether we have the ability to characterize whatever it was that made mammals and dinosaurs able to survive before the asteroid strike.

What seems clear is that what we might call the "intuitive" definition of fitness fares well *except* where chance comes into play. By a frequency account of probability, this is to say that the intuitive notion of fitness as measurable by number of offspring succeeds except where the fact of the matter diverges from Aristotle's "always and for the most part" or, in more modern terminology, where the outcome is far from the limit to which most events converge. An important question is whether *any* definition of a real-world phenomenon or of a concept bearing on the real world can withstand the assault of chance occurrences.

### (1) The difficulty of drawing analogies relevant to fitness

This subsection offers a further argument against the propensity interpretation of fitness. The driving insight is that sometimes fitness is treated as a cause ("That lineage survived because it was fitter than competitor lineages.") while at other times fitness is treated as an effect ("That combination of characteristics should make any individual possessing it fitter than those which possess a different combination in the same environment."). The relationship of natural selection to fitness is then one of *indicator* and *mediator*. Natural selection as a force *indicates* how fit an organism is by the degree to which it hinders what would be (in the absence of any obstacles) the organism's longevity and reproductive accomplishments. Natural selection also *mediates* the fitness of an organism in so far as it makes outwardly manifest something which would otherwise remain hidden. But the two causal relationships in which fitness thus figures are what we might call heterogeneous or asymmetric. That is because fitness as effect is manipulated in part by environmental factors (e.g., weather) which cannot be described in terms of their fitness, while fitness as cause has effects which similarly are not describable as fit or unfit (e.g., when flatulent termites change the composition of the earth's atmosphere by contributing up to 30 percent of its methane content, q.v. n. 32 below). In other words, the pattern is either (1) or (2) below, both of which juxtapose fitness with something which is "a-fit" -- neither fit nor not fit.

(1) An a-fit (non-fit) cause yields fitness through natural selection.

(2) Fitness yields an a-fit effect through natural selection.

We will see that this situation differs from those in which some other forces and quiddities (e.g., gravity and mass) figure. In those situations, there is a certain homogeneity or symmetry: bodies having mass take on the roles of cause and effect with respect to each other's movements through the force of gravity. Both "sides" in these juxtapositions of cause and effect share the same nature -- both can be described in terms of mass -- so that an immediately understandable transitivity is possible. A body having mass affects other bodies having mass, which in turn affect it. To speak of cause is automatically to speak of effect. By contrast, when someone (e.g., Mills and Beatty) treats fitness as an "intrinsic property" of an organism, it is not at all clear what is meant unless further information is given. Do we mean fitness as effect or fitness as cause?

### (a) forces in general versus evolutionary forces

Darwin is renowned primarily for answering a vexing question -- What causes change of such a magnitude that speciation occurs? -- by explaining the concept of natural selection (1859). The basic idea of natural selection is so well known as not to need repeating here. What matters for our purposes is that Darwin does not just cite natural selection as a force, but also specifies the *cause* of the force. To go this next step it was necessary for Darwin to appeal to another entity (environment) external to whatever it is that changes. That is because the force, natural selection, cannot be described solely in terms of the organisms which it affects in the way that other forces in nature -- gravity, say -- can be explained as in some sense originating within the bodies which they affect. Whatever bodies are affected by gravity are also the source of gravity. This is true even though such forces can be named and quantified without mentioning any *particular* bodies. We can talk meaningfully about gravity in the abstract, for example by noting that it is associated with an acceleration of 32 ft/sec$^2$. To put the case in a nutshell, some forces integral to various scientific disciplines not only *act on* bodies of various kinds but also *originate* wholly within the same bodies. The two aspects of such forces -- their effects and their origin -- can be separated conceptually, but the overarching theory asserts an intimate link between force and object.

The same is only partly true of natural selection. Perhaps the best way of demonstrating this is to try drawing analogies. We know that a crucial element in a model of gravitation is, simply, the notion of a body. What would be the analog of this "body" in the theory of evolution? If an analog exists, it clearly must be something physical. That leaves us a fairly wide range of choices, and which of the repertoire we choose may, for the reasons suggested above, depend upon our purpose. But let us stipulate that our immediate aim is to explain the concept of natural selection in very broad terms, ignoring to whatever extent possible the fine points of the units of selection controversy. In this case, when we seek to know what thing natural selection "acts on" in a sense analogous to the "body" which is gravity's object, we will simply choose one of the possibilities. The gene, for instance, would be one correct answer, but the individual or group may also be possibilities. Even something which is not, strictly speaking, a single bit of matter maintaining its identity across time would be a candidate for the thing on which natural selection acts. "It is always possible to talk about the natural selection of a *behaviour pattern* in two ways" (Dawkins 1982: 27; emphasis added).[26] But for the moment let us simply say that the individual is the thing on which natural selection acts. If necessary, we can amend our choice of the unit of selection after making a point about natural selection.

That point is this: natural selection as a motive force does not emerge from the individuals (or any other entities) which it affects. In this sense, natural selection is not just different than, say, gravity; as a force, natural selection is in a wholly different category than gravity. In other words, one cannot write *body : gravity :: individual : natural selection.* The conceptual obstacle blocking such a simple analogy is the notion of environment. Natural selection *qua* force refers to the interaction of environment with individuals (or perhaps genes or some other level). Moreover, environment consists of more than individuals and genes. Of course the environment is composed, even largely composed, of other organisms which are themselves composed of genes. But it also encompasses non-living things, entities not subject to its effects. Obvious examples are climate (including weather, average length of days) and terrain (land or water, elevation, etc.). Less obvious (but related) are hidden or irregularly present agents such as volcanic eruptions and the movement of tectonic plates. Those phenomena are relatively frequent, but events such as large meteor strikes, however rare, may also have played a significant role in extinctions. All such non-organic factors are critical to any model of evolution in general and to any explanation of natural selection in particular. Natural selection's proper analog of gravity's body would have to be something like "individual or the rest of the environment." The interaction of bodies is symmetric in so far as gravity affects them; proportionate to their respective masses, each of two bodies exerts the same force on the other. But this is not true in the case of natural selection as an interaction between environment and individual (or phenotype, gene, group, or whatever). Obviously the environment exerts a force on the organism. The organism also exerts a force on the environment to the extent that the environment includes individual organisms in its composition. But because the environment consists of other factors, ones which remain unaffected by individuals (e.g., the temperature of the sun), there is no sense in which the interaction between a given gene and its environment is symmetric. The interaction *cannot* be symmetric, because a significant portion of the environment is beyond the individual's causal reach. At least this is true at any given moment. One might concoct a science fiction scenario in which human intelligence -- itself the product of individuals or genes -- finds means of manipulating environmental factors which had long remained out of reach. But fairy tales aside, individuals affect parts of but not all of their selective environments, and that is all we need to assert an asymmetry contrasting with the symmetry of a force such as gravity.

### (b) Categories of force

Let us note as an aid to upcoming arguments, then, that there seem to be two broad categories of what could be called "forces." One type of force, including gravity, appeals to the same entities for explanation of cause and effect: gravity is caused by bodies and it also affects the bodies which cause it. Observe any two bodies and gravity as a force can also be observed in the quantitatively regular movement of the bodies (assuming one has capable observational tools -- instruments, in a word -- as well as the information to separate out other forces and the affects of other bodies). The second type of force, for which our paradigm here is natural selection, cannot make the same claim. Assuming a god's-eye view of the development of genes as replicators, they can be perceived as affecting one another in so far as they are conceived as sharing an environment. But they do not by themselves comprise their selective environment, and so, even in theory, natural selection is not demonstrated merely through their mutual presence. Let us call the first category of forces symmetric forces and the second asymmetric forces.

It might be argued that genes (or whatever unit of selection we care to choose) are closely analogous to gravity's bodies, since in fact two genes as replicators do influence each other just as two bodies in a model of gravity do. The fact that there are other things, external to the two genes, which also influence them is true of real-world phenomena involving gravity as well. By this account, a model of natural selection can indeed be constructed by referring to idealized genes, just as it suffices for a model of gravity to feature two bodies. One could simply quantify the relative rates of replication of two genes *qua* replicators (a term frequently used in Dawkins 1976, 1982). Or such a model might consist of a comparison of the respective phenotypes associated with two genes. In such a situation natural selection would be revealed (it might be claimed) by comparing the fitness of each phenotype. But such arguments inevitably fail. They cannot mimic the form of a model of a force such as gravity because the indicators of natural selection -- varying rates of replication, varying levels of fitness, or whatever differential is measured -- are dependent upon a third factor, overall environment. And this is true not just in the "real world," but also on a formal, definitional level. So let us agree that natural selection is not a product of genes *per se* in the way that gravity is a product of bodies.

There is another reason why we should distinguish between symmetric and asymmetric forces, and classify natural selection under the second category. This reason is perhaps more pedagogical or rhetorical than conceptually necessary, but it is important nevertheless. If we conceive of natural selection as emerging from genes, then we lose an important aspect of the theory, one which Sober elucidates by contrasting tennis and golf (1984a: 17). Natural selection as a concept employed in the broader theory of evolution deals with competition, but it is competition in which a living unit of selection such as a gene can "score points" without thereby automatically taking away points from its competitors. Thus natural selection functions in a golf-like rather than tennis-like model. Sometimes the competition is zero-sum, as when winner and loser are diner and dinner, but many times the interaction between living things is much more remote.

It may be possible to draw further distinctions between natural selection and other kinds of forces. First, we can observe that some phenomena which we label forces are *primary* -- they are not seen as the result of other forces working in concert. Gravity seems to be such a primary force. It is evidenced by movement, and certainly one might talk meaningfully about specific components of the overall motion, as when a small portion of a trajectory or orbit is analyzed. But the force of gravity itself, understood as the cause of some instances of motion, is unitary. In the case of an orbit or trajectory, gravity may be just one of the forces responsible for the motion; we may cite another force, perhaps an initial one which moved a body such as a planet into the gravitational field of another body such as a star. But the two forces, gravity and an initiator movement, are conceived as being separable and unitary. Other phenomena can be described in terms of more basic, primary forces. We can call these composite motions *secondary*. It seems that a force such as Brownian motion can be described as secondary in just this sense: as a macro-phenomenon it might be treated as a force, as the cause of certain changes in a given context, but it is itself the product of other, more basic forces causing collisions among particles.

Now perhaps a definitive argument could be made that the force of natural selection is either primary or secondary. The argument asserting its secondary character might appeal to the many ways in which the force is implemented, noting that some of these means could themselves be called forces. A species of bird whose takeoff is too slow may be decimated by fleet predators. Surely gravity plays a role in the takeoff. On the other hand, an advocate for the primary character of natural selection might retort that as force it cannot be reduced to any single more basic, implementing force and so remains unalloyed, i.e., primary. Neither case will be argued at length here. Instead, it will serve us better to understand that natural selection is treated as both primary and secondary in the literature. Often the specific treatment seems not to be a matter of conscious choice on the part of a given author, and indeed many authors waver between the two interpretations in the course of a single book, article or even argument. We will see examples below. The distinction between primary and secondary forces thus becomes important to keep our books straight, so to speak. It matters whether the force is seen in one light or the other. In particular, the meaning and use of the concept of fitness is heavily dependent upon whether natural selection is seen (explicitly or implicitly) as primary or secondary.

The same is true of a further distinction we can draw. It appears that some forces such as gravity are *always* operative among bodies, even though their effects may not always be measurable or, even if measurable, interesting. If the mutual gravitational "pull" between two bodies varies inversely proportionate to the square of the distance between them, then it may be of very little use or interest to speak of the gravitational force existing between Earth and any still-burning star in the Milky Way. But our definition of gravity *qua* force constrains us to assert that its effect exists no matter how small. Moreover, such a force is defined deterministically. Remove other forces from the realm of observation and gravity's effect can be calculated to a precision depending only on the exactness of measurement of two bodies' masses and the distance between their centers of gravity. By contrast, we often speak of phenomena such as Brownian motion as forces, but they are not deterministic in the way that a force such as gravity is. The category of stochastic forces can be understood to encompass forces whose results are -- even in theory -- not predictable in the same way as the force of gravity. Brownian motion may be a bad representative here, since the precise movement of particles may be only practically and not theoretically unpredictable. (A similar distinction between the practically versus the theoretically irreducible character of biological science is the foundation of

Rosenberg 1985). But there do seem to exist truly random forces in nature (q.v. Feynman, 1965, on the "two-hole experiment").

We might also speculate about the temporal aspect of various forces. Presumably some forces act in such a way that their presence and their effect are simultaneous. In other words, one doesn't say, "Well, it's clear that force X is operative in our experimental set-up because we've set up the conditions in which the force is always present. Now we'll wait a while and see what sort of effect the force has." (Of course "natural" conditions would work just as well as a laboratory context in this example; the prerequisite is simply to be certain that a force must be operative, whether or not its effects are evident.) Again taking gravity as our paradigm, it seems clear that we think of the force and its effect as existing more or less simultaneously. Even though there might be a lag between the moment when we know that the force exists and the time when we notice its effect, this in-between period results from an observational deficit rather than a necessary feature of the world integral to the very theory which defines the force. It might be the case that Heidelberg's castle is slowly being toppled by the force of gravity. From this perspective there is a time lag: the force is operative but its effect might not be visible for many years to come, at least when the naked eye is the sole instrument of observation. But no one would doubt that the force of gravity could be observed moment by moment if a few pains were taken. One could simply walk into the castle courtyard and toss a stone in the air. Assuming the stone does not vanish, it is fair to assume that the force of gravity is indeed operative.

Now compare the case of gravity with that of natural selection. If we say that natural selection is operative in a given environment, it is not at all clear that the force can be measured *at any particular moment*. Suppose we speculate that a light-colored moth is at a disadvantage in comparison with a dark one in an industrial region where the trees and other surfaces suitable for landing are stained a darker shade than they would be in the absence of airborne industrial pollutants. How do we assure ourselves that the force is indeed operative? Surely there are ways to do so: no one would deny that over the long term, a statistically significant number of observations of uneven predation could be observed. But our understanding of natural selection is fundamentally different from our concept of gravitation in as much as we cannot test a proposition that natural selection is operative as quickly as we can ensure ourselves that a force such as gravity is in effect. In fact, it seems possible that a phrase such as "is in effect" means something basically different when it is predicated of a force such as gravity as when it is applied to natural selection. When we say that gravity is acting at a given moment, we mean the force is actively affecting the situation *at that moment*. By contrast, the claim that natural selection is the motive force behind the evolution of a given species within a particular environmental context need not, and perhaps cannot, possess the same connotation of immediacy or simultaneity. Drop a stone held at chest height in the castle courtyard and we expect the object to do just that -- drop -- demonstrably and immediately. But place a light-colored moth on a dark tree trunk in an industrial region and there is simply no telling what will happen. A predator might swoop down or might not; the light moth might live to maturity or not; it might have offspring or not. There seems to be no non-arbitrary way of determining whether a badly camouflaged moth's survival is the result of chance or from a lapse in natural selection's agency.

There are many ways one could further characterize the difference between two basic kinds of forces. For our purpose, three major ways will be most productive. The first has to do with the immediacy of falsifiability. The thrown stone falls or it does not. If it does not, we have at least good *prima facie* grounds for claiming that gravity is not operative in that locale. Propositions involving some forces can be falsified (or read "tested") almost at once. This is not the case with natural selection. If a well-camouflaged moth is eaten in the near vicinity of a poorly hidden individual of the same basic type, it seems much more likely that the observer would simply file the incident away among a catalog of observations of predation. The trends which can be inferred from the entire collection then constitute the test of the theory. Because trends (by definition) are revealed only over time, arguably it takes much longer to falsify a claim that natural selection is operative than a proposition that gravity is in effect within a certain context. For want of better terminology, let us say that some forces are *continuous* in their action while others are *staccato*.

A second but related means of characterizing the difference is to observe that statements about certain forces can be tested by observing a *single* object of the agency in question, or perhaps a relatively *small number* of such individuals. Thus we can drop a single rock and feel rather confident that the force of gravity has not gone on holiday. It is possible that we could be momentarily distracted by a barking dog or a strolling mariachi band and that we therefore neglected to observe the stone's trajectory. In such a case we can simply drop another stone. It is hard to set a definite boundary as to the number of trials which would be necessary, but for our purposes it suffices to say that that number is very small (bearing in mind that our focus is not confirmation but rather denial -- seeing a stone sailing upward with no apparent hindrance would mean our theory won't fly). But this is not true in the case of natural selection. We need a huge number of observations to convince ourselves one way or the other, that is, either to confirm or deny that a specific phenotype (or some other unit of observation) is either well or badly adapted within a given environment. Natural selection manifests itself in a staccato fashion, and it's not clear whether the force is actually "there" -- that it's actually operative -- at those times when the light-colored moth sits on the dark tree with utter impunity. Natural selection seems clearly operative at those moments when a bird swoops in an plucks the uncamouflaged moth off the tree bark, ignoring its camouflaged peer. But what if the opposite scenario occurs? What if the bird devours the camouflaged moth and leaves the uncamouflaged one alone? Within the context of a robust (i.e., statistically well-confirmed) theory, one wouldn't say that a proposition such as "The force of natural selection will ensure that, *ceteris paribus,* uncamouflaged individuals become extinct or a minority" is falsified when a single camouflaged moth becomes the entree. This is not to say that there is certainly a sharp distinction between forces such as natural selection and the type of force for which gravity has served as our paradigm. If a stone does rise into the air and vanish, we are more likely to suspect that some sort of trick has been played or that a rare anomaly has occurred, especially if our feet remain firmly planted on the ground. Our impulse would not be to consider that any proposition asserting gravity's agency in the area has been falsified. Nevertheless, it seems one can distinguish between gravity and natural selection as kinds of forces based on how many trials it would take to falsify a given claim, even though no exact dividing line is evident.

Thirdly, we can characterize the difference between natural selection and forces such as gravity by observing that in the case of the latter, the force is identified with its effect. What does gravity do? It pulls things towards a so-called center of gravity associated with every mass; or more succinctly in most mundane contexts, it pulls things down. What is its effect? It pulls things down. Natural selection, on the other hand, is much harder to characterize in such a way that its agency and effect can be seen as similar. We can say that its effect is to ensure that the best adapted organisms will thrive within a given environment from a long-range, statistically defined perspective, and perhaps its agency could be described in the same broad terms: natural selection is that which achieves this goal. But in many propositions natural selection's agency is defined in much different, much more specific terms, ones which are unified only by the abstract formulation of effect. For instance, one might say that the force of natural selection will increase the number of individuals who, for the most part, are well camouflaged with respect to their environments. Caveats can then be added to take into account phenomena such as bright coloration which may be advantageous in attracting and selecting mates among some species. But there is of course much more to natural selection. As a force, its bailiwick is hugely broader than the quality of camouflage. Natural selection acts as a sort of universal gardener, weeding out some genes (or organisms or phenotypes or groups or whatever we understand to be the unit of selection *du jour*) while passing over and thereby preserving others. The range of concrete effects which natural selection can have is vast; thus the definition of natural selection as a force is far removed from any description of one of its effects. By comparison, gravity's abstract range of agency is not so remote from a description of how it acts in any specific instance.

It is interesting to consider how many forces and phenomena are defined in terms which boil down to patterns of motion. Gravity, electromagnetism, centrifugal and centripetal effects -- all are clearly defined in terms of motion with respect to bodies.[27] Similarly, many chemical phenomena are ultimately described in terms of attraction and repulsion. In such cases, the motion is real and simple -- it is to be understood literally as movement and the movement is simple in the sense that the model can break down complex phenomena into relative motion among a very small number of bodies. Perhaps this is an oversimplification with respect to the notion of model *qua* descriptor and predicter of process, particularly in light of the puzzles which physicists and others face. ("...[T]he more you see how strangely Nature behaves, the harder it is to make a model that explains how even the simplest phenomena actually work. So theoretical physics has given up on that" (Feynman 1985: 82)). But models serve not just as descriptions of complex processes;they also figure in working definitions of forces, i.e., of agents which explain the phenomena we observe in some laboratory conditions under some, perhaps coarse standards of measurement. It is this latter sense of model which is intended here. It would be a stretch, to say the least, to describe the effects of natural selection in terms of the motion of existing bodies. From one perspective, the problem is that the units which could be said to move are ephemeral -- there is little temporal continuity even among genes as individuals rather than forms. There is continuity among types, for instance, among the kinds of genes, phenotypes, individuals or groups we observe, but we are less interested in their spatial position than in their numbers. In the end a species may be well adapted regardless of where on the planet its members are. What counts in many contexts is simply how many there are.

It is unclear whether natural selection is deterministic or stochastic *in theory*. For the moment, let us leave that question open. But it seems clear that natural selection as the motive force behind the change depicted in the "tree diagram" of evolution is a heterogeneous force and that it is *practically* stochastic. This latter qualification means that at no point in time can the future be perfectly predicted nor the past perfectly inferred from extant data. Taking natural selection as the primary force operative in evolution, we can guess how a given organism might change if we assume certain things about its future environment. Moreover, looking backward, we can speculate as to how the organism's ancestors might have appeared. But in fact the organisms may evolve differently than we had surmised, and fossil evidence may prove our best guesses about the past to have been false as well.

Perhaps a graphic summary of these distinctions is in order.

| | **natural selection** | **gravity** |
|---|---|---|
| Is the generic source of the force the same as what the force affects (e.g., "body" in the case of gravity)? | No | Yes |
| Is the force "primary" (i.e., *not* a composite of other forces) or "secondary" (i.e., composite)? | Secondary | Primary |
| Is the force deterministic or stochastic? In other words, if the force is in effect, can its effects be theoretically or practically predicted? | Theoretically and practically stochastic. | Theoretically deterministic. Its practical character is an open question. |
| Is the force's agency continuous or staccato/intermittent? | Staccato | Continuous |

| | | |
|---|---|---|
| How many individual instances of the force's agency must turn out other than expected before one can consider a proposition involving the force (e.g., that the force is in effect in a given context) to have been proven false? | More than in the case of gravity | Fewer than in the case of natural selection |
| How strongly is the force identified with the specific changes which it causes? | Weakly* | Strongly** |

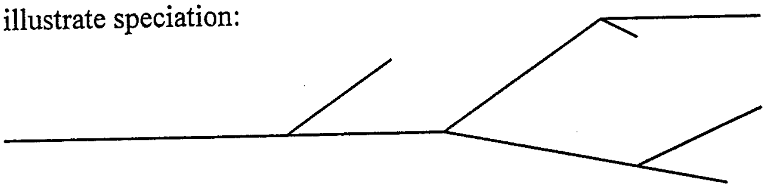
* That is, only abstractly. Natural selection is the long-term weeding-out of poorly adapted characteristics within a given context. It is not the consumption of a single moth which happens to be poorly camouflaged.

** That is, concretely. For instance, in some sense gravity *is* the pulling down of objects.

Where is all this leading.? The point of this list of distinctions can be summed up in two theses. The first is that natural selection as a force is asymmetric, heterogeneous, stochastic and signifies primarily within the theory of evolution. The second is that because of its nature -- described by these qualities -- arguments which appeal to natural selection cannot explain phenomena in what I will call a linear as opposed to a circular fashion. These distinctions are important because they make it difficult to understand how fitness as propensity can be an intrinsic property of an organism in the same way as, say, mass can be a property of a body. Additionally, the conclusions just drawn help to build part of the argument that circularity is a *necessary* feature of evolutionary speculation -- something we will take up in the next chapter. We can call this part of the argument a lemma. Its thesis will be constructed with greater care below, but the reader deserves to know at least the lemma's broad outlines now, while the discussion of the symmetry-asymmetry and deterministic-stochastic distinctions are still fresh. The lemma, then, goes something like this. Evolution intends to explain change in one entity S as a function of a second entity E. (E can stand for "environment," for instance. For reasons which will be made clear later, S is short for "supervener.") Now E contains S, that is, all of the units of selection, whatever they happen to be, are *part* of the environment. But the converse is not true, which is to say that S is a proper subset of E. Thus although we will speak of natural selection as *caused by* E and *acting on* S, natural selection has something of a reflexive character: It is E acting on a part of itself. A moment's reflection should tell us that we could say something very similar even if we dropped S out of the picture. In other words, whatever entities and forces exist in the natural environment, living or not, they affect one another. We could talk coherently about the evolution of weather, for instance.

(2) The random component of repeated change

Consider a simplified version of the kind of tree diagram often used to illustrate speciation:



First we might consider what it means for the tree to represent a model of evolution. Levins notes that "...a satisfactory theory is usually a cluster of models" (1966; Sober 1984c: 27). One other note on the nature of models is in order here, and again we can take Levins as our authority. "The validation of a model," he tells us, "is not that it is 'true' but that it generates good testable hypotheses relevant to important problems" (1966; Sober 1984b: 26 - 27). What kind of testable hypotheses could a "tree model" generate?

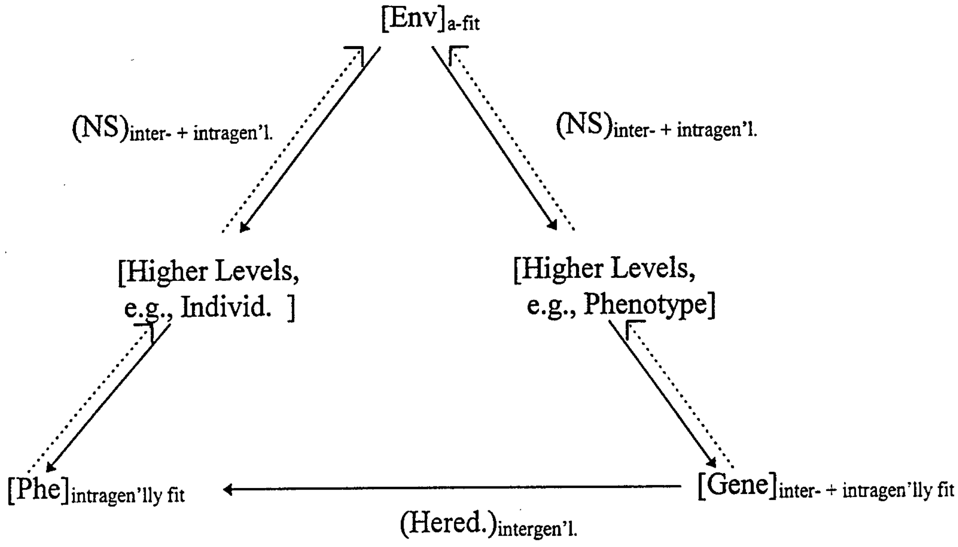
In essence evolutionary "trees" are spatial metaphors suggesting that any given kind of organism has two possible long-term futures, modification or extinction. By modification we understand a continuum punctuated by nodes which signify reproductive separation. This separation can be realized in two ways: either organisms represented by the line on one side of the node cannot reproduce with those on the other side of the node, or else the two "sides" can reproduce but do not except under extraordinary conditions (such as those created in an artificial laboratory setting). The implication seems to be that organisms "blend" into one another by modification of parts. What does not happen under the *usual* conception is for some organism to be incorporated more or less whole into another. To look at it from the other perspective, one organism does not encompass all of another organism. We might make this point more abstractly by saying that an organism is itself; it is not itself-plus-other. But apparently this is neither true of the past history of life on earth and it may be untrue of the future as well. Margulis (1977) suggests that certain organelles in eukaryotes may once have been independent organisms. They entered other organisms as what we would think of as parasites and were then "incorporated" over numerous generations, *merging with* another species rather than merely *emerging from* one.

It is unclear what sort of change the tree represents. For there is a second feature of the picture, as integral to it as change but perhaps less easy to discern. The lines, even when punctuated by vertices signaling the beginning of new lines, represent *continuity*. There is a tension between these two elements -- change and continuity -- akin to the problem of identity which has engaged generations of philosophers. What does it mean to be a single thing, yet a changeable thing? That is one abstract version of the problem. In our present context, the problem could be phrased something like this: What does it mean to say that a group of organisms which appear serially along a time line changes, but sometimes not enough to destroy its "group-ness," its identity as a single thing? Apparently the ancestor-descendant relationship suffices to relate the members to one another no matter how sharply a group member in one generation may differ from a member belonging to another generation. But the individuals themselves die out rather quickly, certainly too quickly for there to be any question of evolution itself being definable by appeal to a given individual. Furthermore, the relationship itself is not a physical thing. Either the identity of the lineage depends on this non-physical thing, relationship, or else there must be a physical entity which (unlike the individual organism) exists continuously across generations. It has been suggested that the gene is such an entity (Dawkins 1976, 1982, 1986, 1995). Of course "gene" is itself a problematic concept. In this case we mean the gene as the unit of selection. Dawkins has suggested calling this sense of gene the "optimon," while Ernst Mayr proposes the term "selecton" (Dawkins 1982: 81). Although this view has its advantages, it further divorces the entity from the force. If what was said above about gravity and "homogenous forces" was correct, then such forces are caused by and affect the same entities or class of entities, namely, bodies. But we have seen that natural selection acts on individuals and is caused by something else, environment. Now yet another element -- the gene -- may be injected, one which allegedly plays an integral role in the process of evolution but which is neither the cause of natural selection nor the thing which natural selection affects. It may not be too great an analogical stretch to say that genes play the role of medium in this case (although they function differently in other areas of biology), almost as though they constitute the ether through which the agency of natural selection is transported.

Relating genes to ether may seem silly. After all, genes exist; ether does not. But there does seem to be a point of comparison if we concentrate on the conceptual role which genes play rather than just their physical reality. Physically, genes are segments of chromosomes. Functionally or conceptually, they are sometimes treated as "information for which there is a favorable or unfavorable selection bias" (Dawkins 1982: 287). What this means in concrete terms is that the environment, acting through natural selection, destroys some phenotypes entirely, allows others to thrive, and exerts a range of influences in between these two extremes. (The positive extreme, which we might call "thriving," is admittedly nebulous, but that is unavoidable at this point in the discussion.) But just as it once was thought that a force must propagate itself through a physical medium, the ether, so we find it unsatisfying to claim that the environment works *directly* on phenotypes. If we were to say that, then the theory of evolution as a whole would have a major point of discontinuity expressible by the obvious question: How are changes heritable? Genes fill this explanatory role, as in the implied definition of adaptation exemplified by the following statement:

> Biological communication is the action on the part of one organism (or cell) that alters the probability pattern of behavior in another organism (or cell) in a fashion adaptive to either one or both participants. By adaptive I mean that the signaling, or the response, or both, have been genetically programmed to some extent by natural selection. (Wilson 1975: 90).

This state of affairs is perhaps best expressible by seeing two related forces and three entities. One can almost envision the relationship as triangular, only note that the gene becomes both the recipient of a force (natural selection) and the cause of a force (heredity). That is, the gene acts and is acted upon. By contrast, the other entities involved are pure with respect to their roles: the environment can be seen as purely active, while the phenotype can be seen as passive. (Of course this is an oversimplification. After all, the environment is itself composed of phenotypes.)



The diagram depicts a scenario in which the environment as an a-fit (neither fit nor non-fit) entity exerts a causal agency on phenotypes and genes through the force of natural selection, mediated through whatever levels are "higher" in a given model. The units of selection problem is obviously relevant here. When we say that we want to know how evolution works, we could just as well say that we want to know how force and entity interact to produce the phenomenon we call evolution. But if we have identified the force (or perhaps only *a* force, possibly one among many) operative in evolution -- namely, natural selection -- we still need to know what sort of entities this force acts upon. Group selection has largely fallen out of favor, making the individual the unit of selection of choice for many researchers. The best-known alternative to individual selection often is referred to by the phrase made popular in Dawkins (1976), "the selfish gene." By this way of reckoning, we start with the proposition that "[e]volution is the external and visible manifestation of the differential survival of alternative *replicators*....Genes are replicators; organisms and groups of organisms are best not regarded as replicators; they are *vehicles* in which the replicators travel about" (Dawkins 1982: 82). Dawkins goes on to define replicator as "anything in the universe of which copies are made. Examples are a DNA molecule, and a sheet of paper that is Xeroxed. Replicators...may be 'active' or 'passive', and, cutting across this classification, they may be 'germ-line' or 'dead-end'..." (1982: 83).

Note the difficulty of applying a non-recursive propensity interpretation to the schema represented by the diagram. We cannot use the forces of natural selection and heredity as a kind of "ether" to move serially from the environment as cause through the gene to the phenotype of higher levels. The reason is two-fold: first, the force of natural selection does not "pass to" the gene without "passing through" the higher levels of quiddity; secondly, in so far as natural selection is held to act continuously on certain levels of life, the progression of forces does not advance serially from natural selection to heredity. Instead, natural selection acts in a kind of circular fashion, with heredity moving in smaller "circles" within the larger, natural selection-driven pattern.

Notice that if a propensity is taken to be an expression of the probable relationship between some level of life and that level's selective environment -- say of individual and environment -- then there is no single propensity even at a single given time. That is because the relationship of environment to individual can be understood as being mediated in various ways (directly or through lower or higher levels), while natural selection can be taken as an intergenerational or intragenerational force, combined with heredity or not. In the next section we underscore the fact that calculations of probability (propensity) are concretely affected by an observer's definition of the entities involved as either unidimensional or layered.

## 5. Combinations and Specific Combinations

In his essay "Are Pictures Really Necessary? The Case of Sewall Wright's 'Adaptive Landscapes,'" Michael Ruse suggests that philosophers have "remarkably little" to say about the ubiquitous phenomenon of biological illustration (1995: 34). In very rough terms, Ruse concludes that biological illustration can be the vehicle not just for transmitting biological information, but also a spur to its creation. Thus pictures such as his paradigm, Wright's pictures of peaks and valleys corresponding to the highs and lows of biological adaptation, play an important and honorable role in biological thought; such pictures are not ill-conceived substitutes for words or equations. That pictures can mislead as well as instruct is a fact Ruse gladly grants. What he does not spend a great deal of time considering is whether pictures can be indeterminate, implying one thing to one scientist but offering a totally different insight to another researcher. If pictures are metaphors or representations of conceptual metaphors, as Ruse suggests, then it would be a surprise if the interpretations of a given picture were all alike. Instead, we would expect scientists to interpret the same illustration differently depending upon their theoretical commitments. (Ruse does an admirable job of showing the other side of the coin, namely how scientists develop different pictures to represent the same idea.)

Ruse seems never to consider whether there is a qualitative difference between words, equations and pictures *qua* symbols. This is somewhat surprising, since a defense of biological illustration clearly could be made by arguing that all science, perhaps all knowledge, is communicated in symbols -- whether verbal, mathematical, pictorial, or whatever -- and that no category can be dismissed as less worthy than the other without justification. To criticize the use of pictures in biology, one would first have to establish a hierarchy of symbols, and then show that, based on some criteria (small likelihood of misinterpretation, for instance), certain categories of symbols (verbal ones, say) are better than other kinds. As I say, Ruse never explores this rhetorical possibility.

But we should. In fact, we should broaden out perspective by asking whether there are some concepts whose exposition is primarily visual, whose interpretations vary, and which are not correctable given other sorts of symbols -- words and mathematical equations, for instance -- that could complement the pictures. Consider Ruse's paradigm, for instance. If a dispute over the proper interpretation of Sewall Wright's adaptive landscapes should arise, presumably the matter could be settled by turning to the long, highly technical paper which the pictures were intended to summarize. (Wright refined the landscapes when he was asked to address a group whose tolerance for long mathematical expositions was highly suspect. In some sense, then, there is an unequivocal truth of the matter -- if not in the pictures themselves then at least in what Wright intended the pictures to portray.) Is there always such an objective arbitrator of what pictures should mean? Perhaps not. Moreover, if pictures are liable to misinterpretation, then maybe the same could be said of other representations. To this point part of our criticism of a propensity interpretation of fitness is that any concrete application of such an interpretation is amenable to various understandings, depending on how an observer defines the entities involved.

To underscore this point, we can turn to a basic problem in probability theory. In the next section, I try to show that the point at issue is directly relevant to reckoning the probable genetic composition of a species where breeding is at least somewhat random. The next section ties the issue of multiple interpretations of numerical results to our general theme of circularity. There I argue that the only way to come up with a stable interpretation of statements about probability is to appeal to convention. The relevant conventions, in turn, inject an element of circularity into the conclusions which are drawn by doing the math.

Before offering the basic numerical example, it is worth emphasizing that what we are talking about is not a very general interpretation of statements about probability. Our concern here is not whether a finite frequency theory is to be preferred to a logical relation theory, for instance. The possibility of multiple interpretations which concerns us here emerges on a different, lower level. Whereas many working scientists would presumably consider their work totally unaffected by arcane distinctions of logical relation versus finite frequency, the kind of phenomenon exemplified below hits much closer to home. A difference in interpretation of first, pictures, and then of the equations which work out the basic visual insight in greater detail, makes a real, significant difference in practical work. The numbers don't match under the interpretations to be offered below, and although there is a "school solution," a canonical way of doing the calculations, there is no clear standard beyond this umbrella of convention.

Among the underlying assumptions of the propensity interpretation of fitness is the notion that we can ask definite questions of the form, "What's a given organism's propensity to have X number of offspring in a given environment?" where each propensity is to be assigned a definite number. Moreover, each such question is thought to have a definite numerical answer, assuming that the environmental variables have been adequately specified. Any vagueness, in other words, must be the result of a lack of information. In turn, that lack can be solved by providing additional (or more precise) empirical data. Now we consider the possibility that even though we may possess all the relevant data, there is still room for a variety of possible answers to questions such as the one posed above. This indeterminacy originates in the way that the question is interpreted. If I simply proposed two ways of evaluating an organism's fitness in some context, it would likely be objected that really the matter is more clear-cut than I make out, since evolutionary biologists "know what they mean" when they pose and solve questions about fitness.

But that is just what I want to emphasize. The matter turns on conventions that are themselves questionable. Thus we will proceed by considering a well-known exercise in probability theory and seeing that specialists who interest themselves in probability questions differ with one another on how specific aspects of certain problems should be handled. The example is to be found in a ubiquitous textbook, one which has seen several editions. The problem generated a years-long debate in the journal *Philosophy of Science*. At issue was the way of interpreting data provided for analysis and questions about that data. To repeat, this problem in probability theory will provide us an allegorical approach to the subject of probability calculations of the kind which Mills and Beatty propose as the basis of a propensity interpretation of fitness.

Here is the problem:

> Remove all cards except aces and kings from a deck, so that only eight cards remain, of which four are aces and four are kings. From this abbreviated deck, deal two cards to a friend. [1] If she looks at her cards and announces (truthfully) that her hand contains an ace, what is the probability that both her cards are aces? [2] If she announces instead that one of her cards is the ace of spades, what is the probability then that both her cards are aces? [3] Are these two probabilities the same? (1990: 471)[28]

Let us represent the modified "deck" of eight cards (the four aces and the four kings) from which two-card hands will be dealt as follows:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $A_s$ | $A_h$ | $A_d$ | $A_c$ | $K_s$ | $K_h$ | $K_d$ | $K_c$ |

This representation simply shows the ace of spades, the ace of hearts, the ace of diamonds, the ace of clubs, the king of spades, etc. Then we can show all possible two-card hands in this way:

| | 1 (X$A_h$) | 2 (X$A_d$ ) | 3 (X $A_c$) | 4 (X $K_s$) | 5 (X $K_h$) | 6 (X $K_d$ ) | 7 (X $K_c$ ) |
|---|---|---|---|---|---|---|---|
| 1 ($A_s$ X) | $A_s A_h$ | $A_s A_d$ | $A_s A_c$ | $A_s K_s$ | $A_s K_h$ | $A_s K_d$ | $A_s K_c$ |
| 2 ($A_h$X) | | $A_h A_d$ | $A_h A_c$ | $A_h K_s$ | $A_h K_h$ | $A_h K_d$ | $A_h K_c$ |
| 3 ($A_d$X) | | | $A_d A_c$ | $A_d K_s$ | $A_d K_h$ | $A_d K_d$ | $A_d K_c$ |
| 4 ($A_c$X) | | | | $A_c K_s$ | $A_c K_h$ | $A_c K_d$ | $A_c K_c$ |
| 5 ($K_s$X) | | | | | $K_s K_h$ | $K_s K_d$ | $K_s K_c$ |
| 6 ($K_h$ X) | | | | | | $K_h K_d$ | $K_h K_c$ |
| 7 ($K_d$X) | | | | | | | $K_d K_c$ |

## (1) Copi's solution

Direct enumeration yields the following values under what I take to be Copi's interpretation of his own problem:

(A.) Total number of possible two-card hands: 28

(B.) Total number of possible two-card hands in which both cards are aces: 6

| | 1 (X$A_h$) | 2 (X$A_d$ ) | 3 (X $A_c$) | 4 (X $K_s$) | 5 (X $K_h$) | 6 (X $K_d$ ) | 7 (X $K_c$ ) |
|---|---|---|---|---|---|---|---|
| 1 ($A_s$ X) | $A_s A_h$ | $A_s A_d$ | $A_s A_c$ | $A_s K_s$ | $A_s K_h$ | $A_s K_d$ | $A_s K_c$ |
| 2 ($A_h$X) | | $A_h A_d$ | $A_h A_c$ | $A_h K_s$ | $A_h K_h$ | $A_h K_d$ | $A_h K_c$ |
| 3 ($A_d$X) | | | $A_d A_c$ | $A_d K_s$ | $A_d K_h$ | $A_d K_d$ | $A_d K_c$ |
| 4 ($A_c$X) | | | | $A_c K_s$ | $A_c K_h$ | $A_c K_d$ | $A_c K_c$ |
| 5 ($K_s$X) | | | | | $K_s K_h$ | $K_s K_d$ | $K_s K_c$ |
| 6 ($K_h$ X) | | | | | | $K_h K_d$ | $K_h K_c$ |
| 7 ($K_d$X) | | | | | | | $K_d K_c$ |

(C.) Total number of possible two-card hands in which at least one card is an ace: 22

(D.) P(two aces | one ace) = 6/22 = 3/11

(E.) Total number of possible two-card hands in which one card is the ace of spades: 7

(F.) Total number of possible two-card hands in which both cards are aces and one of the aces is the ace of spades: 3

(G.) P(two aces | ace of spades) = 3/7

### (2) Rose's interpretation

Rose (1972) reasons that for the purposes of questions (1) and (2) in Copi's problem, knowing that one of the cards is the ace of spades should be the same as knowing that one of the cards is the ace of hearts, or the ace of diamonds, or the ace of clubs. Intuition therefore tells us that the ace of spades has no privileged position in this particular "game" and for these specific questions, and so we can read Copi's second question as demanding the same numerical answer as the first question. Thus Rose would change the value at (G.) above to read P(two aces | ace of spades) = P(two aces | one ace) = 3/11.

It is clear that serious observers of such a scenario can differ widely in their interpretations of even the most basic questions of calculation. The intriguing and paradoxical aspect of this years-long debate over how a specific piece of information should affect our reckoning of probabilities is that an *intuition* (to use Rose's term) of the significance of a specific bit of foreknowledge equates the specific proposition (one of the cards is the ace of spades) with a much more general statement (one of the cards is an ace). It is open to debate whether this intuition is right or wrong. Copi and Cohen, as well as Dale and Goldberg, would argue it is crucial that we pay attention to a datum in the most specific aspect of its character before calculating probabilities. We should not "jump the gun" by attempting to generalize a datum's value before calculation but should instead let the results of the calculation be the grounds for *post facto* generalization.

In observing relative fitnesses, the goal is not (or at least should not be) to claim that an organism's fitness tends to ("probably will" or "has a propensity to") equal a certain quantity. Rather, the fitness of an organism reckoned as a function of its actual longevity or production of offspring should become a data point for generalization about the relation of certain traits to these qualities. We can imagine that the playing cards represent particular fitness values which are associated with specific traits in a given environment, such that "ace organisms" have a propensity to be in the upper register of fitness, whereas "king organisms" tend to fall into a lower register. However, let us also suppose that an organism can have both ace and king traits. If we consider how traits can combine at this phenotypic level, it seems that there is some ambiguity as to how one should answer the analogs of Copi's and

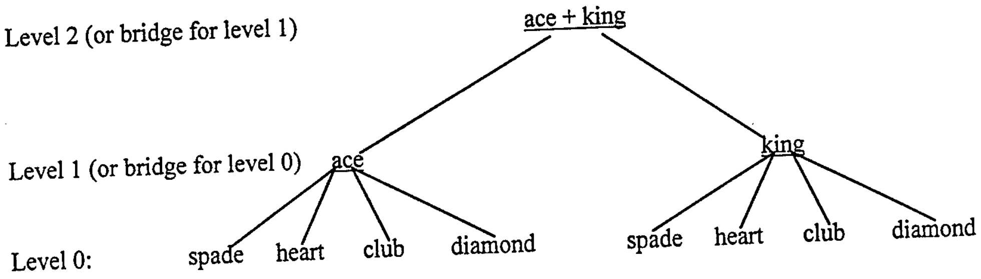
Cohen's questions. If we know that all king-king organisms have fitness value x, for instance, while ace-king traits contribute fitness value y, and ace-ace traits have the higher fitness value z, such that x < y < z, what is the probability that an organism will have fitness value z given that it has at least one ace trait? What is the probability that an organism will have fitness value z given that it has the ace of spades trait?

One could of course write a recursive computer program which would scan the table of possible two-card hands shown above looking for pairs with one ace and then two aces in order to answer Copi's and Cohen's questions as Dale suggests. However, such a routine must already have determined that it is not necessary to distinguish between aces and aces of spades. In particular, if one probabilistically links fitness itself (not just a fitness value, i.e., a number) to a trait, as Sober does in what we have been calling $F_s$, then there is no reason to think of the fitness of a pair having the ace of spades as distinguishable from a pair having any other ace. The recursive interpretation which I have sketched above, on the other hand, avoids linking the definition of fitness with a probability. The probability of a given trait or individual having a certain fitness emerges only *after* the statement involving fitness has been referenced.

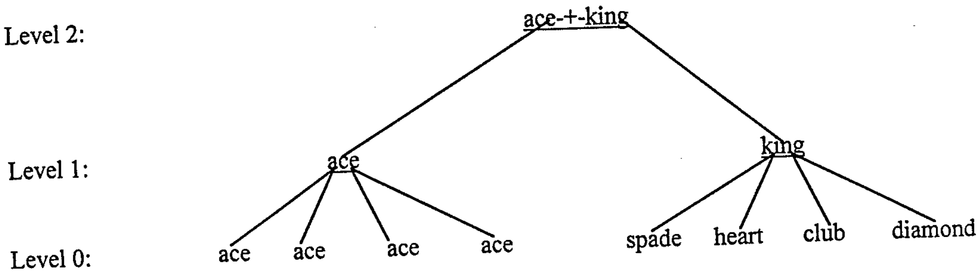
## (3) Other views

Dale (1974) disputes the voice of intuition which Rose hears and insists that Copi's answers (D. and G. above) are correct. Knowing that one of the cards is the ace of spades, insists Dale, differs significantly from knowing merely that one of the cards is an ace. Faber (1976) joins the discussion by suggesting that intuition does lead us to an answer of 3/11 for both questions. Faber reasons differently than Rose, however. Goldberg (1976) essentially endorses what I have offered above as Copi's solution. But in doing so Goldberg differs with Dale on why Copi's answers are correct.

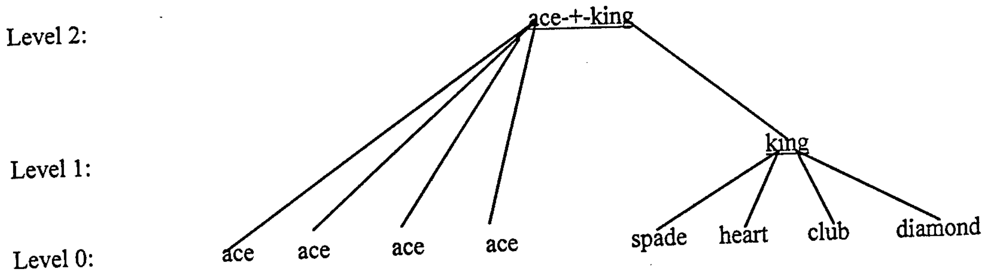
The details of these positions are interesting but need not be rehearsed here. What is important is that the meaning of questions about probability can be answered in various ways without violating any of the basic axioms of probability theory. What happens in the ace-king problem that makes it a problem, that is, liable to varying interpretations? One way of answering is to think in terms of levels of organization. In this case the aces and the kings can be taken as existing at one level, while specifying the suit in addition to the generic denomination takes us to a metaphorically lower level.



Here one can look at all levels except the bottom-most as being "bridges" as well. When one seeks to know how many two-card hands have an ace of spades, one can begin at "spade" under "ace" and, using "ace" as bridge, reach "heart," "club," and "diamond" as species of "ace." Then, using "ace + king" as bridge, one can similarly reach "spade," "heart," "club," and "diamond" under the higher level "king." To find out how many two-card hands consist of two aces, one can "parse" within and across levels in a similar way. Whenever the parsing process begins with an ace, the condition "given that one of the cards is an ace" has been met. (This condition is sufficient but not necessary, since we could begin parsing with one of the four entries underneath the genus "king.") Now given this way of parsing, Rose's claim that "intuition" indicates the conditions "one of the cards is an ace" and "one of the cards is the ace of spades" should yield the same probability seems plainly false. If we were to redraw the diagram treating ace as a higher element corresponding to four identical lower elements, Rose's view becomes more plausible.



In this case there is no need to consider "ace" a higher level at all, since it has no distinct lower levels beneath it. Thus what we have is something like:



But Rose makes another claim, namely, that the condition “one of the cards is the ace of spades” should yield the same probability as if we substituted “one of the cards is the ace of clubs,” “...ace of hearts,” “...ace of diamonds.” This is correct as far as it goes: it does not matter *which* specific identity is given. However, it can matter *that* a specific identity is given. In other words, it makes a difference whether the lowest level specified comprises homogeneous or heterogeneous elements. In saying that specific identity at the level under consideration does not count, we must be very careful not to suggest that the higher level can be eliminated until we know that the present level is homogeneous. And as long as there is a higher level, that is, so long as we require a concept for unifying the heterogeneous members at the lower level, those units must be treated as distinct. That, at any rate, is the viewpoint opposed to Rose's. But is either viewpoint privileged independently of convention? Not so far as I can see.

In the case of a phenotype represented Aa, the issues raised by the various commentators on Copi's ace-king problem seem not to come into play. By a standard representation of how genes combine in diploid organisms, traits come together only within the same realm, so to speak -- roundness and wrinkledness, for instance. One does not consider the problem of a range of dominant traits combining with a range of recessive ones, so that Ab and AB are as much possibilities as Aa and AA. That would be like talking about apples and oranges, as it were, or skipping the bridge at the next level up (in our diagram of the ace-king problem) and using the bridge at a higher level instead.

But the basic confusion about how to count units which we witnessed in the debate over Copi's problem can come into play when we begin to link “levels” of fitness with one another. In particular, it may not be clear what order of levels is proper in making propensity judgments. If a genotypic variation $g_1$ temporally precedes and is associated with a phenotypic variation $p_1$, then it seems proper to call

$g_1$ the cause of $p_1$. But suppose that $p_1$-type organisms -- i.e., organisms possessing $p_1$ as a phenotypic trait -- mate together more frequently than non-$p_1$ organisms, and imagine further that such mating leads to a second widespread genotypic change $g_2$ (heritable change). Then is it proper to say that $p_1$ caused $g_2$? In other words, can we say that genotypic change causes phenotypic change, or vice versa, as a general rule? Or must we leave causality out of our reckoning and simply say that phenotypic change is associated with genotypic change? (And where is the dividing line between the two categories? Is there such a thing as a genotypic change which does not affect the organism's behavior or structure in any way?)

The concept of surpervenience alone does not allow one to answer these questions satisfactorily. We might correctly say that environment can cause a phenotypic change which in turn has no effect on the genetic makeup of the organism under consideration. In that case, it seems clear that genetic makeup can be independent of phenotypic variation. However, it is apparently also the case that environmental factors can change the genetic structure of an individual without causing any phenotypic change in the present or future generations. Mild radiation might cause such an alteration at the genotypic level alone. Thus we have a degree of independence on both sides. Perhaps the argument could be made that *some* phenotypic change supervenes on genotypic variation.

But let us suppose that we can adopt a more or less standard organization, with alleles below combinations of alleles, then perhaps richer combinations which are not recognized as phenotypes, then phenotypes followed by combinations of phenotypes, and finally individuals, groups, and established taxa. Such a representation assumes that variation leading to the modification of existing species and formation of new species, whether this variation results from natural selection or random walk or some other process, is associated with changing combinations of genes and varying frequencies of these combinations. But presumably combining alleles is not simply a matter of asking how many ways a single, anonymous allele can be combined with other single alleles in the case of a diploid organism. This can be seen if we pose the question about combinations in the following way: Given that one of the alleles is $a_i$, how many possible combinations are there? The answer simply is unclear for the same reason that Copi's problem was liable to various interpretations. Resolving to err on the side of safety by always treating heterogeneity as significant (in the way that the difference between the ace of spades and other kinds of aces can be treated as significant or insignificant) in calculating propensity will not rescue us from confusion. We need to know in what sense specific identities are significant.

# Chapter Seven: What is Circularity and Should It be Avoided?

The second major goal of the propensity interpretation of fitness is to avoid what Mills and Beatty term "circularity." This chapter can be viewed as a second major campaign against the propensity interpretation, this time carried out by questioning whether the goal of avoiding circularity is served by a propensity understanding of fitness or not. Evaluating the "second half" of Mills and Beatty's thesis is indeed one goal of this chapter, but further responses to circularity will be addressed as well. Prominent students of fitness other than Mills and Beatty (notably Sober 1984a, 1993) have attempted to defend the theory of evolution as a whole and fitness in particular from allegations of circularity. Here, we defend fitness as well, but by much different, perhaps surprising means. Rather than try to divorce fitness from circularity, we will accept that circularity of a kind seems immune to a propensity interpretation or indeed to any other understanding of fitness. Indeed, if a recursive understanding of fitness is possible, then circularity is not to be shunned; on the contrary, we will presently see that a certain kind of circularity is a help rather than a hindrance to theories of evolution by natural selection and perhaps to other models as well. But first let us explore senses in which the theory of evolution by natural selection can be circular.

## 1. Reviewing the propensity's interpretation's assumptions

In addition to introducing and enlarging a vocabulary of circularity, this chapter will explore what is probably the best known single work in the fitness canon, Susan Mills and John Beatty's "The Propensity Interpretation of Fitness" (1979).[29] As its title suggests, the article does not offer just any perspective but rather presents the *propensity* interpretation as opposed to views relying on something other than propensity. That something could be called *actuality* or *reality*, meaning what *really* happens in a particular case rather than what *tends* to happen in general. Mills and Beatty believed their predecessors had erred by adopting such a reality-based rather than stochastic interpretation of fitness. We can summarize the article's key assumptions as follows.

(1) Assumption 1 (explicit): Evolutionary biologists "agree on how to measure fitness."

Early on in their presentation Mills and Beatty claim that "Biologists agree on how to *measure* fitness" (Mills and Beatty 1979: 264; Sober 1984b: 37). Presumably what the authors have in mind is a set of standard equations which biologists use to measure fitness among various populations. But these equations presuppose a certain heuristic understanding of the concept of fitness. Thus Futuyma prefaces his discussion of such equations with a general definition of fitness: "In an evolutionary context, fitness is measured only by a genotype's rate of increase relative to other genotypes...." (Futuyma 1986: 151). With that basic concept in mind, we can then define the rate of increase r of a genotype in terms of "$l_x$ (the probability of survival to age x)" of organisms possessing that genotype, "$m_x$ (the average fecundity at age x)," and L, the maximum life span of such organisms:

$$\sum_{x=1}^{L} l_x m_x e^{-rx} = 1$$

(ibid.; see Sober 1993: 57-59 for a short derivation of ¯w¯.)

Other equations are used to calculate the rate at which genotypes increase in asexually reproducing populations in which there is no generational overlap and for sexually reproducing organisms.

That there are such equations and that they serve a useful purpose in biology in general seems unquestionable. Furthermore, the general appearance of the equation is *prima facie* evidence that a propensity interpretation of fitness is necessary and therefore correct in at least some respects. Clearly the equation has to do with probabilities and averages (the probability of survival to a certain age and the average fecundity at that age), or perhaps to say the same thing, with propensity. Moreover, the equation obviously conforms to at least one of the axioms of probability theory, namely, P(S) = 1, where S is all the outcomes in the sample space (Ross: 21). Of course the equation must conform to the other axioms as well, but that it does so may not not be obvious at first glance.

But does the existence of such equations imply agreement among biologists on how fitness should be measured? Not necessarily. It is clear that biologists need ways to summarize their own observations and then to compare their results with those of other researchers. And doubtless many use equations such as the one above. But that does not necessarily imply that biologists are satisfied that such measures are ideal nor that these equations measure precisely what should be measured. The fact that there is debate over the unit of selection should be sufficient to tell us that some biologists find equations such as the one above to be less-than-ideal measurements of fitness. The level at which selection is carried out must be the primary level at which fitness is to be reckoned. This issue will be revisited below. For the present it suffices for us to notice that if Mills and Beatty begin by assuming without sufficient warrant that biologists share a common understanding of fitness, the analysis may well fall short of one of the authors' stated goals -- to "provide a propensity interpretation of fitness, which we argue captures the intended reference of this term as it is used by evolutionary theorists" (1979: 263; Sober 1984c: 36).

(2) Assumption 2 (implicit): Allegations that statements about fitness are circular or tautologous amount to the proposition that such statements repeat a *definiendum* in a *definiens.*

A crucial ingredient of Mills and Beatty's argument deals with terminology. Early on the authors make reference to the unfortunate *circularity* which some attribute to statements involving fitness. However, although the term circularity and its derivatives appear numerous times in the article, even in a section heading ("2. The Charge of Circularity") there is no rigorous account of what circularity means in this context.[30] But there are hints. In the introduction we are told that some biological "explanations are no more than re-descriptions of the phenomena to be explained" (Mills and Beatty 1979: 264; Sober 1984c: 37). This approaches the basic notion of circularity as logicians sometimes employ the term: a definition is circular if it repeats the *definiendum* in the *definiens*. But at no point do Mills and Beatty explicitly define circularity in this way. We will tackle this issue again when we consider how Sober deals with tautology.

(3) Assumption 3 (explicit): Fitness is a coherent concept only when it is understood as a propensity of the organism to reproduce.

It may seem odd to label the statement above an assumption instead of a conclusion. After all, Mills and Beatty's purpose is apparently to provide an argument for a propensity interpretation of fitness. But I would argue that their reasoning actually takes this conclusion as an assumption (which is especially ironic in a paper designed in part to combat the allegation that the concept of fitness can only be employed in circular propositions). In fact Mills and Beatty's understanding of the "raw" fitness data which become the indicators of a tendency are themselves frequently mere propensities -- they are not "raw" at all.

In the last chapter we saw two broad ways of reading Mills and Beatty's claim, ways which can provisionally be summarized as follows. First, we might choose to consider an actuality interpretation of fitness to mean that the relative fitnesses of organisms are determined by counting up the number of offspring left by each. In general, the more offspring produced, the greater the fitness. (The article's account is a bit more nuanced in so far as Mills and Beatty recognize that long-range fitness may depend on more than just raw numbers). The second approach is to base fitness judgments on *propensity* to produce a certain number of offspring.

In some contexts a propensity interpretation seems wholly inevitable. Suppose we count up the offspring left by A, then do the same for B, and compare the numbers. For the sake of argument we will assume that there is no need to account for future generations; raw numbers suffice for an actuality interpretation. Then we will say that A is fitter than B if and only if A produced more offspring than B. So far, so good. The trouble is that biologists are more interested in the relative fitnesses of types than in the fitnesses of individuals. But how do we measure the fitness of types according to an actuality interpretation? The only conceivable answer is that we must first formulate taxonomic principles which "create" the type (e.g., taxon) in question, then we must make a large number of observations to check what each type actually does. Now in this situation it can be argued that an actuality interpretation with respect to individuals is inevitably a propensity interpretation with respect to types. That is because we conceive of the type's fitness in terms of a tendency of the fitnesses of individuals.

But how do we delimit the cases in which the propensity interpretation is necessary? What defines such cases? And why does the propensity interpretation prove to be a scientist's (or logician's) salvation in such instances? The key to these questions turns out to be a rather innocuous-looking phrase which turns up several times in the course of Mills and Beatty's article. That phrase is "defined in terms of," and to repeat, we will have to spend a little time deciphering what it might mean to say, for instance, that fitness can and should be *defined in terms of* the propensity to produce a certain number of offspring rather than the actual production of that many descendants. Mills and Beatty understand the process of defining fitness *in terms of* propensity to be a rather abstract business, applicable to all organisms in all environments throughout the theory of evolution just because it is so abstract. The authors consider this to be a huge advantage in comparison with any possible definition of fitness *in terms of* actually realized physical structure or behavior, since these qualities are too context-dependent to be truly general, which is to say universally applicable. Certainly it is to be counted an advantage if a theory -- which is what an interpretation of a concept such as fitness amounts to -- has a broad range of applicability, but we should not accept Mills and Beatty's claim of abstractness without closer scrutiny.

Our discussion of the propensity's interpretation's primary goals -- avoiding logical paradox and escaping allegations of circularity -- has left many questions open. It seems certain, however, that the notion of circularity is difficult to define in such a way that it is avoidable in all instances and for all purposes. In fact, it is not always clear that we should want our scientific propositions to avoid circularity of a specific kind. Recursive functions, for instance, inevitably have a circular character, but that kind of circularity need not equate with empirical vacuity nor with a lack of utility. On the contrary, definition in terms of self can provide elegant solutions to difficult problems.

## 2. The alleged circularity of fitness

Apparently one sort of "circularity," as Mills and Beatty use the term, is *tautology*. Brandon and Sober have pointed our that *propositions* may be called tautologous but that the term cannot properly be applied to a *phrase* such as "survival of the fittest," much less to the single word *fitness* (Brandon, 1978, in Sober 1984c:

65; Sober 1993: 69 ff.; see Sober 1984a: 63 for a less emphatic statement of the same point). That seems quite correct. Sober is also right when he says that many claims made by evolutionists are clearly non-vacant. What is apparently lacking in Sober's account, however, is an explanation of *why* the interpretation of fitness used in the theory of evolution by natural selection seems at times circular -- perhaps tautologous, perhaps analytic, or maybe circular in some other fashion. Perhaps an example will clarify what I mean.

Mark Ridley describes sympatric reproduction in the case of the insects ("lacewings") *Chrysopa carnea* und *C. downesi* (1985: 107 ff.). The case can be summarized as follows. In spring and summer *carnea* is light green; in autumn it is brown. *Downesi* is dark green throughout the entire year. *Carnea* lives in fields and meadows and on deciduous trees; *downesi* is found only in conifers. Geographically, *downesi* lives only where *carnea* is also found. There is no evidence that the species were ever geographically separated from one another. The two species are genetically identical except for three genes. The color (light green/brown or dark green) is determined by one of these three genes. Dark green *downesi's* "color gene" is made up of alleles $G_1G_1$ ; $G_2G_2$ makes *carnea* light green or brown depending on the time of year; $G_1G_2$ never appears in nature.

Researchers C.J. and M.J. Tauber believe *carnea* and *downesi* exist as the result of sympatric speciation. They argue that the two species are not the product of a *geographic* separation. Instead, the cause of the appearance of two separate species where there formerly was only one is the result of a genetic change which occurred within the single geographic range of the original ancestor species. For our present purposes it is important to note that the Taubers' explanation depends on the concept of selective advantage:

> ...There would have been a continual production of inferior heterozygotes of intermediate colour [during the process of sympatric separation], camouflaged neither to conifers nor to non-conifers. A reduction of interbreeding would be advantageous. Any change that reduced interbreeding (and did not cause any other disruption) would be favoured by natural selection. (Ridley 1985: 108)

Ridley does not address what seems to be controversial here, namely the decision to call this case an example of sympatric rather than allopatric speciation. The term allopatric as applied to speciation means a case in which separate species develop because of geographic isolation. In our present example, the two species, *carnea* and *downesi*, share the same *range*, but do they share the same *environment*? Naturally any answer to the question depends on one's definition of environment. Since *carnea* lives only in meadows and on deciduous trees, while *downesi* lives only on conifers, it would seem permissible to consider the environments of the two species to be different even though their range is the same. On the other hand, if environment and range are considered to be the same, then of course the two species share a common environment. In any case, the ancestor species can be said to have occupied a single environment.

We should pay attention to the specific roles which fitness takes on in the speciation scenario hypothesized by the Taubers. Evolution by natural selection (ENS) as a macro-theory for explaining the fossil record as well as the observable diversity of species is, in part, a theory for explaining why there are various species and the mechanism which may have formed them. But ENS does not tell us precisely how species arise. That task requires specific subtheories of speciation. Here the basic question is how it came to be that *carnea* does not breed with *downesi* despite their genetic similarity. Two of the three types in the hypothesized ancestor species (camouflaged for conifers, camouflaged for meadows and deciduous trees) were fitter than the third type (not camouflaged for either "backdrop"). In this way fitness functioned as a *cause* of differential reproduction in so far as the behavior of fit individuals caused that behavior to become the norm among all members of the species (*carnesi* and *downesi* do not interbreed in the wild, thus excluding the possibility of non-camouflaged offspring). In so far as the behavior or individuals exhibiting it can be called fit, fitness is also an effect of the evolutionary process.

This is worth repeating in slightly different terms. An argument that evolution by natural selection can take place requires a theory of speciation. How can a single species split into two? In the case of sympatric separation it is apparently necessary to assume the principle of natural selection. The sympatric theory goes roughly like this: First, a genetic alteration occurs among the members of the single, "parent" species. Those members whose chance genetic makeup causes them to be phenotypically better adapted to their habitat reproduce most effectively. But a single habitat presents more than one selective environment (meadows, fields and deciduous trees on the one hand, and conifers on the other, in the case of *carnea* and *downesi*). The species members who are ill adapted to any of the present environments tend not to survive, while selection favors those who breed only with those who share their very particular adaptive environment (in which case a disadvantageous heterozygote $G_1G_2$ would be less likely to occur). Eventually interbreeding becomes unusual or even totally unknown, causing two or more species to arise from the original one. (*Downesi* and *carnea* breed at different times of the year, again as the result of a genetic difference. However, the insects can be "fooled" into interbreeding in the laboratory.)

The explanation of sympatric speciation fails if one cannot appeal to natural selection, for without this appeal there is no way to account for the disappearance of the disadvantageous heterozygote. Thus a theory of natural selection requiring an explanation for speciation ends up appealing to itself (in a certain sense) within its account of speciation. If we explain an instance of hypothetical speciation by saying that the fit behavioral phenotype is the phenotype which survived (and in turn led to two species in Mayr's sense of two non-interbreeding populations), in what way have we made a circular argument? Have we constructed a tautologous or analytic proposition, or is there some other sort of circularity involved? And regardless of how we eventually characterize the rhetorical pattern, is it logically defective or not? These are the questions we approach in the next sections.

## (1) The alleged analyticity of fitness

In this section and the next we use not only Mills and Beatty's propensity interpretation but also Sober's (1984a, 1993) reflections on tautology and analyticity in order to consider the issue of circularity more broadly. Sober offers the following as a "serviceable definition of *fitness*":

"Trait X is fitter than trait Y if and only if X has a higher probability of survival and/or greater expectation of reproductive success than Y" (1993: 70).

There is apparently no doubt in Sober's mind that this definition (which will be called $F_S$ below) is a tautology, but the characterization does not bother him. Citing his own work (1984a) and that of Kitcher (1982a), he asserts that "[t]he fact that the theory of evolution *contains* this tautology does not show that the whole theory *is* a tautology"

(1993: 70). Thus Sober heaves a sigh of relief, believing that the theory of evolution in general has escaped the allegation that it is tautologous while declining to defend his definition of fitness against this charge.[31]

Throughout the rest of this section, we will proceed as follows. First, we should seek to explain why the definition of fitness which Sober offers (a good one, incidentally) should *not* be deemed a tautology, and then why $F_S$, when viewed as a definition encompassed by the theory of evolution, *does* require defense against the charge that it is a tautology. To demonstrate this latter point, we will review an argument that when a key concept within a system is tautologous, the system as a whole suffers even though, as Sober claims in the case of evolution by natural selection, the system in general may not be tautologous. Next we will consider an explanation of how $F_S$ can be a definition of fitness, as Sober asserts, and *simultaneously* a substantive claim which plays a key role in the theory of evolution by natural selection as a whole. For reasons already offered above, we will avoid appeal to propensity, and we will add to these reasons in the course of the discussion. To repeat, I believe the best defense of fitness explains how the concept functions *recursively* within the theory of evolution by natural selection.

### (a) Why $F_s$ should *not* be called a tautology

Since the word "tautology" is being batted about here, a few words of clarification are in order. Sober himself understands a tautology to be a disjunction of the form "X is P or not-P." He offers two examples: "'It is raining or it is not raining'" and "'Pigs exist or pigs don't exist.'" In explaining why the first of these two examples is a tautology, Sober asserts that "[t]he definitions of the logical terms 'or' and 'not' suffice to guarantee that the proposition is true; we don't have to attend to the nonlogical vocabulary in the sentence (e.g., 'raining')" (1993: 69). Another, more precise definition of tautology can be found in Copi and Cohen: "A statement form that has only true substitution instances is called a *tautologous* statement form, or a *tautology*" (1990: 287). Sober apparently feels that tautologies of the form "X is P or not-P" are not a good tool for discussing the possible tautological character of the definition of fitness $F_S$, since statements about fitness do not have this basic structure. Perhaps $F_S$ could be forced into a disjunction of the form "P or not-P," but Sober seems to think another kind of statement would be a better candidate for demonstrating the logical structure of $F_s$. What sorts of propositions does he think look more like statements about fitness and are at the same time tautologous?

Sober goes on to suggest that *analytic* statements -- such as "All bachelors are unmarried," for which the "logical words" do not suffice to guarantee truth but which are necessarily true on semantic grounds -- are sometimes called tautologies. It is presumably in this sense, the sense of being analytic, that Sober calls the definition of fitness $F_s$ "tautologous." That is, he thinks $F_s$ is tautologous on semantic grounds. (Even though this violates the rigorous definition of tautology employed by logicians, for the sake of argument we can accept a loose sense of "semantic tautology" equating to analyticity.)

Let us return to Sober's paradigm of an analytic proposition to see if he is correct. We recognize the analyticity of the assertion "All bachelors are unmarried" if and only if we know the meanings of the key semantic entities, namely, "bachelor" and "unmarried." Moreover, we must know these meanings *prior* to reading or hearing the proposition in order for it to be analytic. If we had no such foreknowledge, the proposition would provide us with a substantive revelation -- it could be part of a revelatory definition for "bachelor," in fact. (The analogy in the case of tautologies should also be clear: to recognize the tautologous character of "It is raining or not raining," we must understand the logical function of "or" and "not" before we read or hear the proposition.) Thus in order for the definition $F_s$ to be "semantically tautologous," i.e., analytic, we would have to know *before* reading the definition $F_s$ that "fitter" means "higher probability of survival and/or greater expectation of reproductive success." If by "definition" we understand the explanation of a word whose meaning we do not already know, $F_s$ *qua* definition cannot be semantically tautologous (analytic). This sense of definition is commonly referred to as *lexical* or as a *dictionary definition* because it provides an explanation of an unknown word in terms whose meanings are already familiar to us. To make what follows as clear as possible, I will use the term *revelatory* definition in order to emphasize the salient feature of such a definition for present purposes -- the revelation of an unknown term's meaning.

We might also want to read $F_s$ as a stipulative definition, that is, as establishing a convention by which we limit the possible meanings of a term which is already familiar to us. In other words, it might be objected that definitions are of interest within theories even when we know the meanings of all terms on both sides of the copula. A definition might then be not a revelation of meaning, but rather something else -- a way of keeping our theoretical books straight, so to speak, or of ensuring the completeness and clarity of a schema. Let us agree to call non-revelatory definitions -- definitions affirming meanings which we already know -- "bookkeeping" definitions. In that case, a bookkeeping definition could indeed be called analytic, but only *improperly* so. *Improper* is here used in the sense given it by mathematical set theory. (Recall that {1, 2} is a *proper* subset of {1, 2, 3}, whereas {1, 2, 3} is an *improper* subset of {1, 2, 3}.) This distinction between proper and improper subsets shows why it may be ill advised to call even a bookkeeping definition analytic. A definition tells us the necessary relation between a *definiendum* and a complete *definiens*; a proper analytic statement such as "All bachelors are unmarried," on the other hand, informs us about a relation between a *definiendum* and *part of* its *definiens* (between a term and a *proper* subset of its meaning, that is). Thus "All bachelors are unmarried" is not a definition of bachelor because "unmarried (thing)" is merely a necessary but not a sufficient part of the meaning of bachelor. Depending on who is doing the defining, a true definition of bachelor would have to include the stipulations that this unmarried entity is human, is male, is not a widower, is not divorced, etc. If we say a bachelor is any one of these things, or a conjunction of some (but not all) of them, and provided that we already know the meanings of bachelor and these terms, then we have made a properly analytic statement.

A revelatory definition, to summarize, is a mapping of a *definiendum* onto the total *definiens*, and this mapping tells us something which we did not already know. A bookkeeping definition may simply repeat some meaning of which we were already aware, but like a revelatory definition, the bookkeeping definition relates a *definiendum* to its complete *definiens*. A properly analytic statement, by contrast, is a mapping of a *definiendum* onto a part of its *definiens*. Note that a definition which repeats lots of individual facts which we already knew -- e.g., that a bachelor is (among other things) male, that a bachelor is (among other things) unmarried -- can still be revelatory if it is identified as an exhaustive definition. This is because we may have known all the individual associations without being aware that they, taken together, were not just necessary but were also sufficient to define the concept.

Graphically, then, a definition looks like Figure 1 below. We want to know what the term D means, and so we are told a definition, that is, we are given an exhaustive conjunction of features which as a group are to be associated in an identity relation with D. We should of course already understand the meanings of the elements in the conjunction which we call the *definiens*. Otherwise we would have to extend the diagram as in Figure 2.

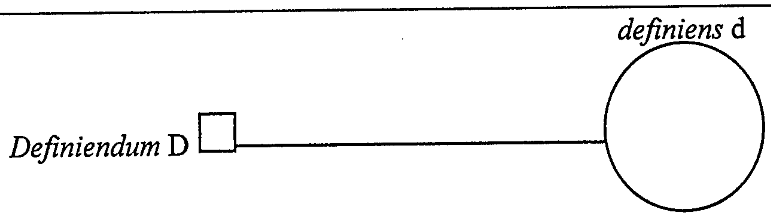


Figure 1: Definition (given every point in d)

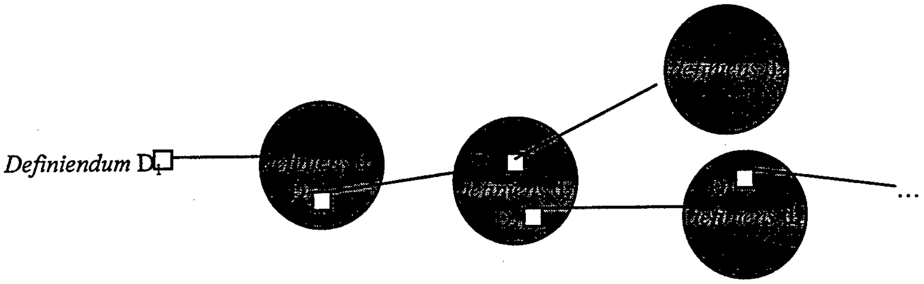


Figure 2: Definition of $D_1$ in which $D_2$, $D_3$, etc. are unknown

In the case of a proper analytic statement (Figure 3), on the other hand, we are offered a proper subset of elements conjoined in the *definiens*, but *we already know the meaning of the definiendum D as well.* This last qualification is important to distinguish between revelatory definitions and improper analytic statements; without it, no such distinction appears to be possible, for an analytic statement in which an improper subset is mapped onto the *definiendum* would then be the same as a revelatory definition.

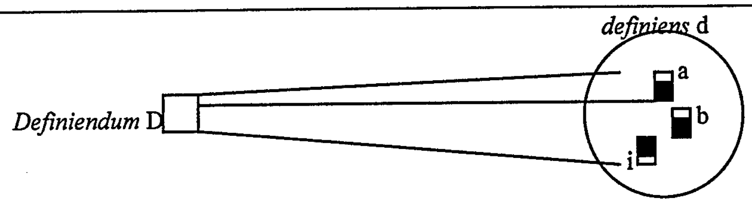


Figure 3: Analytic Statement (given D and at least points a, b, and i in d)

Perhaps the best strategy is not to quibble over whether a definition such as Sober's $F_s$ is analytic or not, but rather to identify different kinds of analytic statements. In the case of "All bachelors are unmarried," it seems clear that there is nothing to be gained by testing the proposition. We could make such a test, and its results could be expressed in something close to "85 bachelors surveyed...85 unmarried." It would be unthinkable for us to obtain any other result, unless an error is made in taking the survey. That is, any deviation from a perfect correlation must be epistemological; no real, metaphysical divergence of the two numbers is possible. The point here is that there is a perfect one-to-one correspondence between the status of bachelorhood and the status of being unmarried. But in the case of fitness nothing like that one-to-one correspondence is possible. This is a rough restatement of Brandon's criterion of independence (1978).

Now let us apply these reflections to Sober's definition of fitness, $F_{s.}$ If $F_s$ cannot be brought into the form "X is P or not-P", then it is not logically tautologous. (Sober seems to suggest this, but he is never explicit in this regard.) If $F_s$ is indeed a definition, as Sober claims, then it must be either a "revelatory" definition (i.e., one from which we learn for the first time the meaning of the *definiendum*) or else it is a "bookkeeping" definition. If $F_s$ is a revelatory definition, we cannot meet a necessary epistemological condition of an analytic statement -- foreknowledge of the complete meaning of the term to the left of the copula -- and so $F_s$ cannot be semantically tautologous (analytic). On the other hand, we are ill advised to call a "bookkeeping" definition tautologous in a semantic sense, for to do so blurs a useful distinction between definition and analytic proposition which mirrors the mathematical distinction between improper and proper subset. In short, Sober should not call his definition of fitness, $F_s$, tautologous, nor is the definition analytic.

For clarity's sake, let me make the same point in a slightly different fashion. Suppose that three mathematicians, Joe, Bob and Fred, repair to their favorite watering hole Friday night after work and there continue to discuss a mathematical object which they have been imagining. This object is called a "saur," short for *sauros*, because it is so large a quantity. By the time the third round of beers arrives at the table, all three have agreed that the saur has properties x and y. Shortly thereafter, Fred excuses himself for a few minutes. Joe and Bob continue talking. During Fred's absence they prove to their satisfaction that a saur must also possess property z. Their definition of a saur is therefore "that which is x, y and z." Fred returns to the table. If Joe recounts the definition for Fred's benefit, Bob hears it as a "bookkeeping" definition whereas for Fred the definition is revelatory. If Joe had announced that a saur possesses property x or y, the statement would have been analytic for both Bob and Fred, whereas the statement that a saur possesses property z would be analytic for Bob but not for Fred.

For the reasons stated Sober can rightly call $F_s$ a tautology only if we assume that (1) analytic statements can indeed be called tautologous in a semantic sense; (2) the recipient of the definition already knows the meanings of all its terms (i.e., $F_s$ is a "bookkeeping" rather than "revelatory" definition); and (3) we allow for "improper" analytic statements. Passing through this set of wickets leaves $F_s$ a very anemic sort of tautology at best. In fact, it seems that Sober would do better not to apply the term tautology to $F_s$ at all, since any more robust senses of tautology simply cannot apply. For the sake of argument, however, let us ignore this problem and suppose that $F_s$ is a tautology, as Sober claims. In this latter case, is Sober right to withdraw from the fray, taking comfort in his assertion that although the theory of evolution by natural selection contains this tautology, the system itself is non-tautologous in so far as it generates other, interesting propositions? To put the issue more abstractly, if a theory depends on claims $\{s_1, s_2,\ldots,s_n\}$, some subset of which consists of tautologies, can the theory as a whole be non-tautologous? That is the question to be considered in the next section.

## (b) Why it matters whether the definition of fitness is a tautology or not

Sober himself says that the description "tautologous" properly applies only to statements. And presumably a scientific theory can in general not be expressed in a single statement. In particular, we would not wish to claim that any contemporary version of the theory of evolution by natural selection can be encapsulated by a statement summarizing all of its aspects. Sober offers a conditional statement to express the way in which

> Darwin was able to characterize how the force of natural selection works in a remarkably simple way: if the organisms in a population that possess one characteristic (call it F) are better able to survive and reproduce that the organisms with the alternative characteristic (not-F), and if F and not-F are passed from parent to offspring, then the proportion of individuals with characteristic F will increase. (1984a: 27)

But this conditional does not summarize the entire theory of evolution by natural selection. If we cannot summarize this theory in a statement, and if only statements can be tautologies, it seems to be unenlightening (or even analytic) to claim that the theory is non-tautologous even though it contains the alleged tautology $F_s$. Perhaps, then, we should not even question whether the theory of evolution is tautologous just because it contains $F_s$. A more salient question is whether the theory -- or any scientific theory, for that matter -- loses explanatory or predictive power if it includes tautologous claims. This question seems unanswerable; to formulate a response, we would have to know what the theory as a whole seeks to account for and what other definitions and claims are part of the theory. But if it is agreed that we should exclude revelatory definitions from the category "tautology," as the previous section recommends, this much seems clear: a theory containing a tautology is less elegant, less parsimonious, than one which includes no tautology among the claims it encompasses. A tautology (which, to repeat yet again, should not be a definition) is itself deadweight. Its logical necessity is already given by the background knowledge which the theoretician possesses. If it is a semantic tautology in Sober's sense -- that is, an analytic statement -- then any meaning it holds is already known to the theoretician, since the meanings of the logical and semantic terms and their relationships to one another must already be known.

To clarify this point, which deals essentially with aesthetics, let $T_1$ be a theory consisting of claims $S:\{s_1, s_2,...,s_n\}$ as well as revelatory definitions D:{d1, d2,...,dr}.

By the convention explained in the previous section, we refrain from calling any of the elements of D tautologous. However, let a subset of S, $S'$:$\{s_i, s_{i+1},\ldots,s_{i+m}\}$, consist of tautologies. If these tautologies are of the logical variety, then they do not contribute to the theory itself. As Sober points out, a logical tautology is a proposition whose truth is evident from the "logical words" it includes. But these logical terms must be part of the background knowledge common to the theory as a whole. Thus a logical tautology provides no useful content at all to the theory, neither in terms of a relation between semantically substantive terms (since the truth of the relation follows automatically from the "logical words" used) nor in terms of the logical words themselves (since their significance and use is already known). Similarly, a semantic tautology (i.e., an analytic proposition) does not offer the theory any information not already contained in the background knowledge or in the set D. In short (not taking into account background knowledge), $S \cup D \cup S' = S \cup D$ because $S \cup D \subset S'$ (i.e., the common elements already contain everything in S').

Here is how our argument against viewing $F_s$ as a tautology stands to this point. First, Sober offers a good definition of fitness (what we have called $F_s$), but this definition, *qua* definition, should not be called tautologous. Even if we stipulate for rhetorical purposes that $F_s$ is a tautology, we cannot take comfort in the fact that a theory which contains the statement is not therefore also tautologous. The theory *is* flawed if it contains a tautology. Perhaps this flaw is "merely" aesthetic, but it seems to me that such a needless addition of useless baggage would nonetheless be a flaw. The alternative, of course, is that $F_s$ is not a tautology. But is $F_s$ a mere definition, or does it serve some other purpose within the theory of evolution by natural selection as a whole? As we discuss circularity and recursion in more depth below, I will try to show how $F_s$ may be understood both as a definition and as a claim which is empirically verifiable.

## (2) Circularity in general

### (a) What do Mills and Beatty mean by "circularity"?

Mills and Beatty begin by reminding us that some have criticized the theory of evolution as essentially empty while others have defended it. It is interesting to observe that the authors believe there is a fundamental problem in evolution, and that this problem causes the ongoing debates. The main problem, according to Mills and Beatty, is that the meaning of fitness is unclear.[32] The perspective of fitness which is often unwittingly adopted (so say the authors) leads to "circularity."

But what do Mills and Beatty mean by "circular"? What do they think that those who level the allegation against statements involving fitness mean by the term? Unlike Sober, who methodically considers "tautological" and "analytic" as possible synonyms for "circular" (1984a, 1993), Mills and Beatty offer us only indirect glimpses of what they take the meaning of "circular" to be.

> "Where fitness is *defined in terms of* survival and reproductive success, to say that type A is fitter than type B is just to say that type A is leaving a higher average number of offspring than type B. Clearly, we cannot say that the difference in fitness of A and B explains the difference in actual offspring contribution of A and B, when fitness is *defined in terms of* actual reproductive success" (1979: 265; my emphasis).

The key phrase here is "defined in terms of," and we should consider very carefully what this phrase, repeated twice in short order, actually means. It seems clear that if fitness is a property applied to individuals or traits, as Mills and Beatty believe[33], then the property must precede its effect, which is alleged by those whom the authors criticize to be the actual or average contribution of offspring to succeeding generations. (It is worth noting that Mills and Beatty sometimes treat the view they are considering as depending on the average actual number of offspring, as in the citation above, while elsewhere they leave the modifier "average" out of the picture. This is significant because of the close relationship between an average and a propensity.) To say that fitness is "defined in terms of" offspring contribution is, in the abstract, to claim that the *definiendum* in this case is defined in terms of its effect, or perhaps one of its effects. In other words, the *definiens* describes the *definiendum* indirectly by describing an effect. This is indeed odd. Let us consider for a moment *definienda* which, like fitness, are alleged to be qualities of organisms. We could narrow the list down to relative terms like "big" and "small" (what Dawkins calls "fading out" qualities). We might even make this more precise by saying that fitness belongs to a genus of qualities which are "Cambridge-changeable." (A Cambridge change occurs, e.g., when a qualitative superlative can no longer be predicated of an object because of something which happens not to the object in question, but to its environment. Thus Bill may be the tallest person at a cocktail party at ten o'clock in the evening, but a moment later he is no longer the tallest. It is not that Bill suddenly became a dwarf, but rather that someone taller has entered the room.) Perhaps we could take weight or "heaviness" as our paradigm for the moment. Like fitness in Mills and Beatty's conception, heaviness is one quality of organisms. And like fitness, it can be defined according to its effect within a given environment. That environment must be considered stable, if only theoretically, lest a Cambridge change occur and counteract the effect of the quality. An organism may leave fewer or more offspring behind depending on the environment, and its weight may also change depending on the force of gravity in the environment. Thus heaviness is a property of the individual in one sense, but to put the matter more precisely, heaviness is something which describes one facet of an organism's interaction with its environment. That interaction *can* be described in terms of an effect. For instance, it is possible that if we were to observe the day-to-day activities of another human being for a year, we would see that the subject occasionally steps onto a scale -- perhaps in the morning after stepping out of the shower, after working out in the gym, or as part of a physical checkup at the doctor's office. From these "natural" events, we can observe the organism's weight. But this does not mean that the needle pointing to a gradation on a bathroom scale is the subject's weight; the measurement is only the effect associated with a certain instance of that weight.

Can we apply this analogy to fitness? Apparently there is nothing to stop us. According to Mills and Beatty fitness can be seen as a property of an individual, but just as in the case of heaviness, it seems more appropriate to view fitness as just one aspect of the interaction between an organism and an environment. A change in either the organism or the environment can change the organism's fitness. For that reason, it seems inappropriate to attribute fitness only to the organism despite the convenience of using the possessive attributing fitness only to the organism (the "organism's fitness"). The analogy with heaviness goes further. Although we can measure fitness by an effect (usually offspring production or longevity), those measurements *qua* effects within an experimental or natural setup are not themselves the phenomenon which causes them within that context.

Now there are a couple of questions which we should consider in light of Mills and Beatty's theses. First, do these authors treat fitness as a property of the individual rather than an aspect of the individual's interaction with an environment? The answer is unclear. Although they say explicitly that fitness is a property of the individual (or pheno- or genotype; see above), at other points in the paper they acknowledge the role of environment in fitness. Secondly, do Mills and Beatty confuse fitness with its effect? In other words, do they identify an effect of fitness such as production of offspring or longevity with fitness itself? The authors seem to attribute this position to those whom they criticize as identifying fitness with the actual production of offspring. However, the way in which Mills and Beatty assert that fitness is a propensity of organisms to produce a certain number of offspring in a given environment seems to amount to the same confusion. To see this, let us consider what we can mean by "propensity." The term can refer to the source (the wellspring) of a given effect, as in the statement, "These birds have a propensity to fly south at this time of year." On the other hand, the same word can be understood to refer to effect rather than cause: "The propensity of the birds is to fly south." In this second sense it is not clear whether propensity means the act itself or the cause of the act. One can only decide which is meant by the context. Now in the case of Mills and Beatty's thesis, it is clear that propensity is equated with an effect, albeit stochastically. (Note that Beatty and Finsen, 1989, question what a propensity is in the sense of inquiring what kind of function it is in a given case -- e.g., arithmetic versus geometric mean -- but they seem not to have considered the more basic issue of whether fitness functions as cause or effect or perhaps both in a given scenario.)

Mills and Beatty claim that their propensity interpretation of fitness reflects the way in which practicing evolutionary biologists actually use the term. Paying attention to how a concept is employed in a scientific discipline is a laudable goal, and in fact it may advance our own investigation toward a different conclusion than Mills and Beatty's. Is it not reasonable to assume that when an evolutionary biologist looks at a given feature of an organism -- long, powerful legs in the case of a land-dwelling herbivore, say -- the relative merit of the feature may be appraised only indirectly in terms of longevity or production of offspring? Perhaps fitness can be *defined in terms* of (to use Mills and Beatty's phrase) a constellation of characteristics relative to a given environment, a constellation which is too broad and complex in its elements and the interactions among them to describe exhaustively. If that were the case, then one could parallel the practical irreducibility of fitness to qualities with the practical irreducibility of biological science in general (Rosenberg 1985). But theoretically, it seems possible to define fitness in terms of interrelated characteristics within an environmental context. Never mind for the moment that the synergy of such characteristics is so hugely complex that they cannot be described. For our theoretical speculations of the moment, what counts is that such characteristics *must* exist. That is, there must be an optimal solution or solutions to the "engineering " challenge facing an organism *once one or more qualities are specified.* For instance, if it is stipulated that an organism lives in a semitropical environment and is a herbivore which feeds in daylight, presumably it will be advantageous to have a quality such as sharp eyesight. The reason why we cannot say that it is good for an organism, that the organism is at its fittest if it possesses characteristics which we could name as superlatives -- sharpest eyesight, fastest, hardest to see, hardest to smell, etc. -- is that the resources which an organism possesses are limited. That means that if an organism is fastest, it may also need to be smaller than some of its potential competitors, as cheetahs are faster but smaller than and therefore vulnerable to lions. This line of reasoning is reminiscent of Kitchener's evaluation of the dinosaur *baryonyx*. (Recall that the argument was based on what we called "hypersufficiency" -- a too-perfect conjunction of traits -- and asserted that evolution by natural selection operates by providing adequate rather than "overkill" solutions to the challenges posed by various ecological niches.)

Thus Mills and Beatty make a valid point when they suggest that it is impossible to come up with an all-purpose list of qualities which belong to the fittest organisms. Fitness is a subtle concept, highly context-dependent, and it cannot be equated with or (again to use Mills and Beatty's phrase) *defined in terms of* specific qualities. On the other hand, is it not self-evident that fitness is to be equated with leaving the maximum number of offspring?

Mills and Beatty's answer to this question seems to be positive, while the correct answer to the question is negative. (Beatty and Finsen 1989 is essentially a record of the authors' epiphany in this regard.) The problem with trying to define fitness in terms of, for instance, "the fastest" or "keenest sighted" is that superlative in one sense need not equate with optimality. But the same is true of the number of offspring produced. Moreover -- and this is a difficult obstacle for the propensity interpretation -- in this respect it does not matter whether one speaks of real or probable numbers of offspring produced, that is, of particular reality or general propensity. Examples demonstrating the non-optimality of maximum offspring production are easy to find in the literature. Among field mice in an area of eastern Canada, for instance, it has been demonstrated that it is more likely that *more* of the members of the largest observed litter will die before reaching reproductive age than in the case of somewhat smaller litters. Thus if one looks at survivability of a *lineage* rather than merely at the number of offspring which constitute a given generation at a particular time, the superlative, "most numerous," is not the same as the optimum in the sense of "most survivable" (Morris 1986: 174, 178).

Do Mills and Beatty manage to avoid the circularity which they attribute to statements which equate fitness with actual number of offspring produced? Apparently not. As has been shown above, the circularity which they identify is a kind of equivalence: in the case of an actuality interpretation of fitness they hold the *definiendum* and the *definiens* to be equivalent to one another, and therefore the juxtaposition of one with the other results in an empirically vacuous claim. But that kind of circularity results from the equation of fitness -- however the term is understood -- with an effect of fitness, or in other words with a measurement of it.

### (b) Sober's approach

We have used Sober's 1993 perspective on tautologies and analytic statements to investigate his $F_s$. Sober tends to avoid the term "circularity" (which Mills and Beatty (1979) use with such frequency), but based on the discussion above one can imagine how this visual metaphor might apply to an analytic statement. Because the predicate is already contained within the subject of what we have called a *proper* analytic statement, the statement as a whole has a self-referential quality: it juxtaposes a set S with a proper subset of S, as though to emphasize the aspect of the subject which is named by the predicate. Thus in Sober's example, "All bachelors are unmarried," we ignore aspects of bachelor such as *human*, *male*, etc., and focus on the one aspect *unmarried*. Diagrammatically, one might represent the circular aspect of an analytic statement by showing that there are parts of the statement which are identical (e.g., *unmarried*) -- even though one has to look, so to speak, *inside* the subject to recognize this -- and that these identical parts "point to" each other.

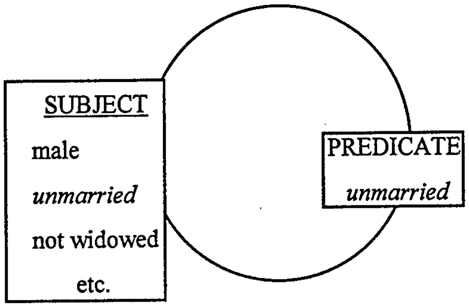


Figure 6: Circular nature of a proper analytic statement

Now in this diagram it is not clear *how* the recipient of an analytic definition, whether of *bachelor* or *fitness*, perceives the rhetorical quality which we metaphorically call circularity. To take the case of fitness: if Sober's definition of fitness (which we have called $F_S$) is taken as revelatory, then apparently the recipient would have known in advance what the *definiens* means. The semantic and logical units (if any) of the predicate are already known, and the information which is imparted is the association of those known elements, in the form in which they are concatenated in the predicate, with the unknown *definiendum*, fitness. One might therefore say that the circle goes from what is known in the direction of a new association, or in other words from the *definiens* to the *definiendum*.

In the case of a bookkeeping definition, by contrast, the recipient already possesses a conception of the *definiendum* as well as knowledge of the content of the *definiens* prior to encountering the specific association represented in the analytic statement $F_S$. It is therefore not clear what the direction of the metaphorical circle would be in such a case. Perhaps it is a matter of emphasis, or of the direction (so to speak) of a question one wishes to answer in an organized fashion. A backward-looking question based on comparing the fossil record with modern life forms, for instance, might seek to know why one of two closely related ancient varieties of a reptile species became extinct while the other variety continues to flourish in the present day. Assuming that natural selection was the dominant operative force and that $F_S$ is a serviceable bookkeeping definition of fitness, we can answer that the now extinct variety was *less fit* than its peers. That is, we look from the fact of the survival of one variety and the extinction of another, assume that longevity and reproduction are key to fitness in accordance with $F_s$, further assume that greater fitness leads to higher probability of survival, and then conclude that the surviving variety was fitter than the now extinct one, at least in the environments in which the two most recently lived.

Alternatively, we might look forward into the future rather than backward into history, perhaps considering two presently existing varieties of a species and asking which of the two will probably flourish to the greatest degree in the future. We may already have compiled "track records" representing the respective longevities and reproductive rates of the two varieties within what we predict will continue to be a stable environment. In such a case, are we looking from the predicate toward the subject, that is, from the *definiens* toward the *definiendum*, or in the opposite direction? Presumably we would not decide the matter by claiming that one variety has proven fitter than the other unless we had already measured the substantive terms in the *definiens* of $F_s$, namely respective longevities and reproductive rates. If that is the case, then the *definiens* is in some sense prior to the *definiendum*, as in the case of the backward-looking question posed above.

## 3. On the impossibility of avoiding circularity in some contexts

The primary claim of this chapter is that circularity -- of a kind -- is *necessary* to propositions used as part of a hypothetico-deductive method. If the argument is successful, then the "fitness problem" in evolutionary biology will evaporate, or at least part of it will. Obviously this position is a far cry from strategies which flee from allegations that fitness is circular, either by attempting to modify the understanding of fitness (as Mills and Beatty do with their 1979 propensity interpretation of fitness) or by surrounding a *sua culpa* with evidence that evolutionary biology makes some non-circular claims even if not all propositions involving fitness are informative (e.g., Sober 1984a, 1993).

Roughly, the circularity problem consists of the charge that the foundational logic of evolutionary theory is flawed because of a "circularity" contained in propositions such as "Among organisms and species, only the fit survive." In its most basic form, the criticism suggests that the phrase "fit organism or species" has no meaning within the theory other than "an organism or species which will survive." If this were true, then the proposition above would amount to the empirically empty claim that *organisms or species which will survive will survive.*

Evolutionary biologists and philosophers of science who are inclined to play Defender of the Faith have taken two basic strategies in dealing with this fitness problem. One group digs in immediately and attempts to wrestle a non-circular conception of fitness from the literature since Darwin, thus showing that there is a legitimate (i.e., empirically meaningful) way to interpret propositions which associate fitness and survival. The second group flees from propositions such as the one above while fortifying other claims of evolutionary biology. By this account, the linkage of fitness and survival in propositions is unenlightening, but that is an accident of the scientific and philosophical dialogue, an accident which does not affect the legitimacy of the field of evolutionary biology as a whole.

The kind of benign or progressive circularity I have in mind occurs when the test of specific hypotheses of the same form as the general hypothesis are understood to indicate the truth or falsity of that general hypothesis. In other words, the hypothesis that "the fittest organisms survive" is tested when we seek to determine whether some specific organism survives and whether that organism is judged fit based on previously inferred criteria. What I will *not* do below is defend the value of specific tests by divorcing them from some general statement or statements. This latter strategy is adopted by Dawkins when he discusses the possibility that certain wasps who fight over possession of burrows in which they have invested part of their personal resources conduct themselves in accordance with the "Concorde Fallacy." This principle dictates that the value of resources already expended rather than the evaluation of the real future worth of something determine whether an individual continues a course of action or not. The namesake is the Concorde aircraft, of which Dawkins says: "one of the arguments in favour of continuing with the half completed project was retrospective: 'We have already spent so much on it that we cannot back out now'" (1982: 48). He adds:

> If we were interested in testing the general hypothesis that animals optimize, this kind of *post hoc* rationalization would be suspect. By *post hoc* modification of the details of the hypothesis, one is bound to find a version which fits the facts. Maynard

> Smith's (1978b) reply to this kind of criticism is very relevant: '...in testing a model we are *not* testing the general proposition that nature optimizes, but the specific hypotheses about constraints, optimization criteria, and heredity'. In the present case [regarding the wasps] we are making a general assumption that nature does optimize within constraints, and testing particular models of what those constraints might be. [Dawkins 1982: 49]

There is no question that specific propositions *can* be tested. But in fact a general hypothesis encompassing those specifics *is* being tested at the same time. If a specific proposition of the form of the general hypothesis fails, then that is *prima facie* (but not always sufficient) evidence that the general hypothesis is false, though it may be salvageable with modification. If a specific proposition is found to be true, on the other hand, then the general hypothesis which generated it is strengthened to a degree.

### (1) Inferring fitness in circular fashion

Suppose that we are not trying to avoid the first problem confronted by Mills and Beatty's propensity interpretation -- pardoxes stemming from chance occurrences -- but rather are striving to escape the charge of circularity. In such a case, can we appeal to anything other than propensity? The answer would seem to be affirmative if we consider induction in a certain light. The argument, offered in greater detail below, goes like this. If we are presented with a certain phenomenon, we are entitled to formulate a hypothesis as to how that phenomenon occurred. Our hypothesis may include an overarching theory which will also predict how future events related to the phenomenon will play out.

An explanation of an existing circumstance may take the form of a story. If the story makes sense, then the hypothesis can be accepted on the condition that no contradictions be revealed in the future. But the only sensible way to tell some such explanatory stories is to weave the hypothesis into the narrative. If the hypothesis is that Colonel Mustard was shot in the library by Mrs. Green, then we might assume the hypothesis and try to find some conceivable scenario under which the hypothesis would fail. Maybe Mrs. Green has an alibi -- perhaps she was in another city at the time of the murder, for example. Perhaps she is blind and therefore probably could not have used a handgun to commit the crime. If there is no way of contradicting the hypothesis within the narrative, then we have some (possibly weak) *prima facie* evidence for its verity.

Mills and Beatty (1979), Sober (1984a, 1993) and others have suggested that fitness can be understood probabilistically as a propensity toward a certain degree of reproductive success or perhaps longevity. As we have seen, Mills and Beatty in particular propose that fitness is a sort of quality inhering in an organism which causes a certain probability of reproductive success. In turn, this quality is associated with that organism's fitness, perhaps even to the extent of being equated with it. Sober has used the phrase "dispositional property" to apply to probabilistic as well as deterministic properties of objects which, when in the presence of a "triggering condition," are revealed in certain probabilistic or deterministic behaviors (Sober 1984a: esp. 47). Applying such ideas to the concept of fitness in evolutionary biology, Sober as well as Mills and Beatty imply that fitness takes its meaning from a property or properties inherent in an organism.

There is a problem with this view, a difficulty which stems from the fact that fitness is essentially a *comparative* matter, one that talks about the *respective* longevities and rates of reproduction of organisms *within a given environment*. Thus a given organism can remain stable in all of its structural and behavioral aspects, yet its *fitness* can change because its environment changes. Moreover, even if the organism and its non-organic environment remain constant, the organism's relative fitness can change if its fellow organisms change in some way. So broadly speaking we have three possibilities under the heading "fitness change." (1) First, an organism O in a stable non-organic environment E can change in its makeup or behavior while its peers remain unchanged. In this case, O's fitness understood as some sort of value -- perhaps the mean of its expected numbers of offspring -- may change. But this case is of interest only when we take into account how O's fitness changes with respect to O's peers. (2) This modication of fitness values can also occur in a second case, one in which O's fitness remains constant in a stable non-organic environment but in which the average fitness of O's peers changes. (3) And of course it is possible that O and its peers remain unchanged while their environment changes.

These latter two cases, in which O's fitness changes even though O remains constant in its constitution and behavior, might be termed Cambridge changes. In case (2), what we might call O's "absolute fitness"--that is, O's expected longevity or its expected number of offspring -- remains constant, but its "relative fitness" changes. In other words, O is more or less fit with respect to its peers because their mean fitness has changed. Case (3), in which O's environment changes, depicts a change in O's absolute fitness even though O itself has not changed. Thus the interesting aspect of an organism's fitness -- not its absolute fitness in a world held artificially constant, but rather its fitness relative to its environment and its peers -- is not a quality which remains stable. That is not an earth-shaking conclusion, of course. More interesting is the fact that the quality's comparative significance can change even though the quality remains constant in an absolute sense. The organism's particular makeup and behavior affect its fitness, but aspects of the world outside the organism, the organism's peers and environment, play as great or greater a role. A devotee of the propensity account might well object that the theory is perfectly intelligible if one simply stipulates that these other factors must be held constant. But why should one hold such factors to be constant? In some of the most interesting and dramatic aspects of the theory of evolution as a whole, changes in environment and peers' fitness are obviously critical. The theory that dinosaurs became extinct because of meteor impacts causing an abnormally opaque atmosphere and therefore colder-than-normal temperatures is an obvious case in point. Even if we do hold the factors of environment and peers' fitness constant, it should be obvious that an individual organism's fitness is of interest only in so far as it is understood as a function of these other factors as well as of qualities of the individual itself.

It is noteworthy that Mills and Beatty begin their discussion of the concept of fitness with the explanation of what they term "$\text{fitness}_1$." This sort of fitness is the source from which the concept of "relative $\text{fitness}_2$" is derived. In this respect the organization of the argument seems not to mirror what one would consider the "natural" development of the concepts, at least with respect to what we will call the "forward-looking question" of Darwinian evolution. "The $\text{fitness}_1$ of an organism x in an environment E" is defined as "the expected number of offspring which x in E will leave behind" (in Sober 1984c: 46).

But why would anyone be interested in such an isolated number in the first place? Darwinian evolution asks two basic types of questions, one historical and one predictive. The historical one, which we will call the "backward-looking question," asks why organisms of a certain kind are seen at a given point in history. Why are those organisms in particular present? Why did they survive while, as is clear from the assumptions of the theory as well as the fossil record, other organisms became extinct? Of course there is the possibility that factors other than natural selection -- random walk or chance cataclysms -- may have played a role. But insofar as natural selection is assumed to have operated, the backward-looking question seeks to flesh out the obvious and in itself uninformative answer ("The surviving organisms were fitter in their environment") by investigating what fitness means in this context.

The goal of such interest must originally be construction of a method to compare the relative reproductive rates (real or tendential) among various organisms within a taxon. That is, the Darwinian theory of evolution asks as perhaps its most basic *forward-looking* question: "Which of the organisms is fitter than the others?" That is the organism which *will* (or will probably) live longest and leave behind the most offspring. Perhaps an analogy will help to make this point more intelligible. Consider the words "size" and "big" and "small." An object O can have a determinate size, represented perhaps by a single integer. A one-dimensional object, for instance, might be of size seven units. Now for an object of its class, seven units might be big, small, or just about normal. One cannot say much about O's bigness or smallness until one knows more about O's peers and their sizes, and perhaps as well about O's world in general. A flea might be of normal size and yet still be considered small. O's absolute size -- seven units -- is a property of O itself. But we cannot properly label O "big" or "small" or "or normal size" based only on whatever intrinsic properties O possesses. In fact, even O's absolute size of seven units is well nigh meaningless without reference to some other facts about the world, perhaps facts about other objects of O's sort or other aspects of the world at large. A Cambridge change is always possible when O is viewed in context. That is, it is always a possibility that at one instant we may judge O to be big, while because of the entrance of new objects or simply on the basis of a different perspective, O's seven units seem later to be rather insignificant. We can of course stipulate that we wish to "freeze" the context of observation by stipulating something like "in environment Ei at time Ti," but that alone does not lessen the role of comparison in making O's size meaningful. For suppose that we make such a stipulation about O's environment and peers, and then take a measurement and find O to be seven units in size. What meaning can such an observation possibly have unless we already have in mind a list of general facts about things in the world which have a size of seven units and a list of particular facts about objects of O's type which are 7 units in size? What we want to know is whether the size is big, small, or normal, or associated with some other interesting property. The relative perspective is the one that matters in evolution by natural selection, and the relative aspect is at least once removed from absolute quantity.

The analogy with "fitness" and "fit" should be clear. An organism O may have a certain absolute fitness defined as an expected number of offspring, say, in a given environment. And that may be, as Mills and Beatty suggest, an intrinsic property and therefore a propensity of the object in the artificially fixed environment. But that does not tell us whether the organism is fit. Fit, like big and small in our analogy, does not take its meaning intrinsically, but rather comparatively. And that fact is crucial in evolutionary biology, where an organism's fitness often becomes especially interesting when it changes. As we have seen, such interesting changes can take place in Cambridge fashion, that is, because some new factor enters the context of observation. Fitness can indeed be viewed as an intrinsic propensity, but that perspective accounts for the least interesting (i.e., non-comparative) aspect of the concept. The soul of the concept is relative and not intrinsic to a single organism or taxon.

One can call a key concept such as fitness comparative in a rather narrow sense or in a much broader one. The narrow sense is the one which we have just reviewed, in which fitness is likened to the relational terms *big* and *small.* One need not commit to any particular philosophy of science or of scientific explanation to accept this view, and if one rejects it, one can remain neutral with respect to issues such as theory-ladenness. But fitness and perhaps other concepts in evolutionary biology may be understood as relational and non-absolute in a broader, *a fortiori* sense in accordance with a commitment to a philosophy of science such as Toulmin's (1972). By this account virtually no concept can be abstractly defined independently of cultural and theoretical commitments.

The history of philosophy contains many admonishments aimed at correcting deficits in human knowledge. Among them is Bertrand Russell's advice: "The first step towards philosophy consists in becoming aware of these defects ... in order to substitute an amended kind of knowledge which shall be *tentative*, *precise* and *self-consistent.*" The defects he means are the "cocksure, vague and self-contradictory" character of "knowledge in ordinary life" (1927: 1 - 2; my italics). It is doubtful that modern readers will want to buy all of Russell's 1927 vintage epistemology, influenced as it was by behaviorism and positivism. But there may be something worthwhile in the constellation of attributes which he claims for ideal philosophical knowledge.

This value is in the apparent oddity of the association. On Russell's view, to be tentative is a necessary (though not sufficient) condition of knowledge which is precise and consistent, but why should those qualities go together with tentativeness? Why can't knowledge be "cocksure" and still be precise and consistent? The answer to these questions requires us to consider more closely what it means for knowledge to be tentative. Presumably tentative knowledge is a provisional conclusion about something worth knowing -- the causal relationship existing between two natural phenomena, say. To be tentative or provisional in such a context would mean that the object of knowledge such as a proposed causal relationship is doubted. It might be as we think it is, but then again it might not. Our uncertainty would cause us to revisit the issue over and over, and that would entail looking at all relevant aspects of the relationship. We would need to study the two phenomena and the environment in which they occur in as much detail as possible. Perhaps we would also wish to test our conclusions experimentally, that is, by abstracting what we take to be the key aspects of the phenomena and their relationship and then reproducing them in a laboratory setting.

Two things seem to be clear in such a situation. First, tentative knowledge in Russell's sense will reference similar sets or iterations of data. This is because an item of tentative knowledge can be *stated* in a proposition, but it can only be *justified* by appealing to the data directly supporting it. (Of course the rationale for believing a proposition depends upon logic and what we might call the scientific "background" of theories in addition to specific data.) In so far as the same phenomena are observed over and over, data supporting the proposition will be gathered in iterations yielding more or less the same results each time, assuming the theory is stable. In other words, because roughly the same phenomena are repeatedly revisited, the information gleaned each time the phenomenon is re-evaluated may well look much the same as what was collected before and what will be collected in future observations and experiments. If such uniformity is lacking, the proposition or the methods by which the data supporting the proposition were collected, or both, will be called into question. Thus each iteration of data gathered in association with some item of tentative knowledge is a test not just of a proposition or theory but of all the other data-gathering iterations. Put another way, researchers expect not just consistency between theory and each iteration of data, but also between iterations of data themselves. Iterations of data should converge to a norm over time, or else the theory will be modified.

Second, it seems clear that tentative knowledge as Russell understands the concept is necessarily theory-laden. In order that the phenomena underlying a tentative proposition can be revisited, they must first be identified. Thus, as mentioned above, a process of abstraction must take place: researchers must decide which phenomena are relevant to the proposition under consideration while discarding other possible objects of inquiry as irrelevant, if only temporarily. (This is not the place for a lengthy discussion of what "relevance" means. Instead, we can accept for the moment something like Popper's criterion of falsifiability as the arbitrator: if a phenomenon might serve to falsify our tentative proposition, then it is relevant and must be studied. But someone will have to decide which phenomena might play this role of spoiler, of falsifier, and making that decision requires further theoretical commitments.)

Concrete examples illustrating these points are plentiful. Taking theory-ladenness first, we can immediately find examples in our current area of interest, evolutionary biology. Belief or disbelief in the possibility of descent with modification, i.e., of evolution, obviously plays a role in how we choose to interpret phenomena available to us. Bones discovered in the Neander Valley near Düsseldorf in 1856 (three years before publication of Darwin's *Origin*) were at first not thought to belong to an extinct race of hominids who may even have been among our ancestors. Instead, the unusual skeletal structure was taken to be the remains of some deformed human being -- an individual different from the norm of his fellow humans, certainly, but belonging to the same species. The subsequent discovery of so many similar bones would have made this view hard to hold onto, but it was the theoretical commitment to evolution, or at least to its possibility, that allowed researchers to identify Neandertal Man (Pfeiffer 1969: 160 - 162). Following that theory-based identification, newly discovered bones and artifacts were taken as further evidence for a theory of human evolution, but that theory needed to be assumed -- made an hypothesis -- in this process of support. This is an accepted, hardly noticeable kind of circularity. In the case of Neandertal, we can postulate that one of the mechanisms which most theories of evolution hold to be operative, natural selection, is indeed a potent force. Then we can explain why extant bones appear as they do and why a species under consideration seems to have existed as long as it did. The theory of evolution holds natural selection to be an important mechanism driving periods of change and stability. Our hypothesis consistent with this commitment is, in a phrase, that Neandertal man was relatively well adapted. We check the facts and find them to be consistent: large cranial size and various artifacts indicate intelligence, which is obviously a survival advantage. Grave sites and communal dwellings indicate sociability, which we can assume made hunting more efficient, helped eliminate starvation as a cause of mortality, aided self defense, etc. Stockiness made for strength and efficient retention of heat in an ice-age climate. (Recall that a similar pattern of reasoning accounted for the sympatric speciation of *Chysopa downesi* and *carnea.*)

Now let us try to make these two basic aspects of tentative knowledge -- the iterative and theory-laden character -- more concrete by diagramming them. To diagram is to adopt a visual metaphor, but what visual patterns should we choose? To revisit the same phenomena over and over seems to be, in a word, cyclic. The process repeats its own broad aspects over and over. The picture or shape which corresponds most naturally to a cyclic phenomenon seems obvious: a circle. The progress of investigation traces the same path over and over, although the data collected may be different. That might suggest a spiral more than a circle, but at least it seems clear that the motion is in some way circular, metaphorically speaking.

What about the theory-laden nature of tentative knowledge? This case is more complicated and correspondingly more difficult to diagram. It seems clear that any given observation may be part of the evidence upon which the theory was based. As in the case of the Neandertals mentioned above, it is also true that a claim in a theory must sometimes be tentatively assumed in order that a given datum can be taken as confirmation or contradiction of the theory as a whole. A skull in itself is no evidence that evolution has taken place; to be meaningful, the skull must be integrated in a certain hypothetico-deductive system. Under such a system the skull and the context in which it is observed can be described in a proposition which in turn can be checked for consistency, and if consistent, the skull constitutes evidence of a kind. The "movement" of analysis goes in both directions: from observations which form the basis of the theory, and from the theory which "flavors" the observation.

Admittedly this account is very loose, but it is meant as nothing more than a heuristic indicator that a pattern of analysis which is metaphorically "circular" can also be scientifically legitimate. We take it as a given that the integrity and consistency of science depend in part on the scientist's commitment to the tentativeness of propositions. If something proves inconsistent, it is thrown out or modified. The major rhetorical goal here has been to equate circularity *of a certain kind* with tentativeness -- with the sort Russell meant, a provisionality which is logically consistent rather than otherwise. The argument has been that *tentative* knowledge of a phenomenon (e.g., a thing, a process, a concept) is that kind of knowledge which requires the knower to revisit the phenomenon again and again to test the provisional conclusion. Circularity is an apt visual metaphor for this recurring revisitation. (Perhaps the strength and consequent utility of our visual capacity relative to our other senses makes it tempting for us, as human beings, to describe our ideas in terms of visual metaphors.)

But if we assume for the moment that one legitimate sense of circularity is simply a brand of provisionality, then to call a statement or an argument circular will not necessarily be to allege a logical error. Rather, a certain type of circularity will be seen as an admirable tentativeness and also (if we buy Russell's association) as possessing a greater tendency to self-consistency than non-circular kinds of knowledge. Although we can treat circularity in the abstract, our special interest here is the allegation that the concept of fitness in evolutionary biology is circular in a sense something like Russell's. We have seen that a first step toward accepting circular fitness as a useful rather than a vacuous or even damaging concept is to understand "knowledge" as a process rather than as a condition. *Tentative* knowledge stands in relation to its object as a perennial visitor: the object is examined, then reexamined, then examined again, *ad infinitum*, or until an exit condition (no matter how arbitrary) is provided. Because knowledge of a thing includes knowledge of its context, this background, too, must be investigated and re-investigated. It is premature at this point to leap to a full treatment of fitness, but it is worth reminding ourselves in a preliminary way that fitness is a function of the organism as well as the environment (context). There is more than one way in which knowledge in the abstract can be construed as process -- when the object itself changes, either in itself or in relation to its context, for instance, or when the object is stable but cannot be appreciated in its entirety at a singe glance. We should consider these two possibilities in greater detail.

It sounds nearly trivial to claim that in so far as the "object" of analysis changes, there is no way in which knowledge can be had without revisiting that object numerous times. Perhaps it is a matter of conscious selection on the part of an observer whether the object of knowledge is stable or dynamic. For instance, one might choose to study a historical phenomenon -- the etymology of a word, say -- either as a "thing" or else as a "process." In other words, we can frame a question such as, "What was the proper pronunciation of a given word in a given region at a certain moment in time?" This question seems to correspond to a stable *thing*. But we might also ask a question of this type: "How did pronunciation of that word *change* over time?" Such a question seems obviously to deal with process. Now let us consider a claim which will require some justification: *all* inquiries into principles or causes must inevitably deal with processes. There must be *movement* between a principle and its consequence. To follow that movement conceptually is to pursue an object or objects of study through numerous instantiations and contexts. To substantiate the claim that all attempts to delineate causes follow this pattern would of course require more justification, but for now let us move ahead to a model of how the observer can attempt to analyze a dynamic subject.

There may be many such means or, to put it another way, we could perhaps describe the means in many ways. But there is one particular pattern of analysis which is of special interest for our present purposes because we are interested in circularity. This pattern of analysis is recursion, which we investigated in chapter three above. The term is doubtless best known in the context of computer science. Recall that there its familiar meaning amounts to self-reference: a procedure "calls" (refers to) itself. There is usually an exit condition in a recursive procedure of this kind, so that the circular "motion" of entry into the procedure and then reentry as the procedure calls itself does not go on endlessly. However, it is conceivable that a programmer might interrupt the process manually rather than inserting a built-in exit condition.

If it is true that the propensity interpretation has limited success in achieving its intended goals of neutralizing chance and escaping circularity, then perhaps the theory of evolution by natural selection faces a serious threat. That would seem to be the conclusion of those who leveled allegations of circularity in the first place but of many propensity theorists as well. Researchers such as Mills and Beatty, and Sober, after all, apparently considered the allegations serious enough to warrant a rather elaborate response. How, then, should committed evolutionists respond to the threats to the propensity theory discussed above? As a reminder, in this section our goal is to develop an inductive-recursive account of how fitness should be understood within the overall framework of evolution by natural selection. I hope to show that in a peculiar sense the attempt to deny circularity was misguided. A better approach is to harness circularity as a potent tool in explaining how evolution by natural selection can function. One way of doing this is to legitimize induction and recursion -- argue that they are productive techniques of scientific investigation which nonetheless are circular in a sense.

If it is true, as suggested above, that we can choose an object of analysis to be either dynamic or stable, then we should consider whether knowledge of a stable object can have a legitimately circular character. We have already seen that knowledge of a *dynamic* object can be circular in a legitimate, even necessary, way. We considered circularity as recursion, as a kind of self-reference in which an observer asks over and over again the same basic questions and takes as the possibly changing answer to these questions the dynamic data collected every time the same object of analysis is revisited. But this may seem an almost trivial contention: it is easy to understand why a changing subject must be revisited. A much more challenging case involves static objects. What if we choose our experimental backdrop so that there is no such fluidity, so that the object of analysis remains stable? Would there be any need to see knowledge as necessarily circular in such a scenario?

One obvious answer to this question is perhaps too easy. We might choose to rule out one of the premises, namely, that it is possible to find an object of analysis which in fact is unchanging. We have already noted that evolutionary analysis regards not just a single object but in fact the context of that object as well, since it is impossible to know everything about an object in isolation. Put another way, we can learn about an object by observing its interaction with its context, its environment.

Now environment *per se* has no given boundaries. Let us say that we wish to know everything about the birds which nest under the eaves of the house on the corner. It may well be that these creatures react to what happens in and to the house itself. If the occupants blare techno music late at night or if the roof goes untended so that water leaks on the nest and its hatchlings, the birds may decide to move. But the house is by no means the only environmental factor. The cement factory at the southern edge of town may leave pollutants in the air and the soil, pollutants which affect the birds' food supply. And perhaps the food supply is also affected by the growing hole in the earth's ozone layer, a hole which some researchers believe is threatening frog populations. Presumably the birds themselves will change no matter what happens in the environment: they will be heavier on some days than on others; sometimes they will be asleep and sometimes awake; sometimes in the nest and sometimes on the wing; and they will age every moment, no matter how subtly, and eventually die. In short, there may be no way to choose a truly stable, unchanging object of analysis.

But suppose we acknowledge this and instead aim at choosing an object of analysis which is *sufficiently* stable for a given length of time and for a given purpose of analysis. Then surely the question as to whether circularity is a necessary facet of knowledge is a serious one and cannot be dismissed by simply claiming that all objects of analysis are dynamic. How do we answer such a question? Again in the affirmative. There is presumably no fail-safe means of knowing that we have apprehended all of the relevant features of an object of analysis, even though we have delimited that object in such a way that it is conceptually stable. If we choose our object of analysis (as in the example above) to be the pronunciation of a given word in a certain region at a specific point in time, we still cannot know that we have gathered all of the relevant data nor that we have analyzed those data correctly. Rather, we are always obliged to leave the possibility open that we have made some sort of error, whether of misperceiving, misunderstanding, or of failing to perceive a relevant phenomenon altogether.

## 4. From circularity to recursion

To make the case that circularity can be innocuous or even productive in a theory, we can consider the situation in which an umbrella theory depends on subtheories and vice versa. In this kind of circularity the umbrella theory requires that a certain process take place; this process, in turn, is explained by a sub-theory which assumes the umbrella theory. But for the sub-theory to make sense, it must assume the umbrella theory. Thus there is a kind of circular (i.e., self-referential or recursive) pattern: the umbrella theory assumes itself in a once-removed fashion, through its sub-theory. Here is an example:

Umbrella theory (UT): Natural selection (NS) causes organisms to evolve.

UT$_1$: First translation of the UT: Force NS causes later generations of a lineage to exhibit different heritable phenotypic characteristics (or the same characteristics in different proportions) relative to earlier generations.

Convention: When differences in heritable characteristics among generations of a lineage are sufficiently large (measured according to a standard which cannot be abstractly specified because it depends on the purposes and subjective judgment of an observer), we say that speciation has occurred.

First Assumption (A$_1$): Natural selection as a force exists when environmental conditions influence the development and survival of organisms.

Second Assumption (A$_2$): Sometimes speciation does occur. (At this point we need not be more specific by, e.g., adopting a certain philosophy of speciation.)

UT$_2$: Second translation of the UT (given the convention and assumptions): Environmental conditions cause speciation.

In order to satisfy ourselves that UT$_1$ or UT$_2$ is correct, we need a further theory, one which describes how and under what conditions speciation occurs. This leads us to the following two sub-theories.

First sub-theory (ST$_1$): Speciation can occur allopatrically. That is, a subgroup of a homogeneous population can become geographically separated from the rest of the population. Because of this geographical separation, the two groups (which just after the separation could have been conceived as complementary subsets of the same reproductive population) fall under the influence of differing environmental conditions. This observation suffices, since we know from UT$_2$ that environmental conditions can cause speciation. (Notice the circularity: UT$_1$ and UT$_2$ appeal to ST$_1$, which in turn appeals to UT$_2$.)

Second sub-theory ($ST_2$): Speciation can occur sympatrically. This occurs when a given environment perceived as a range contains different sub-environments for which different alleles possible in a certain species are differently adapted. We assume that the sub-environments can be taken as environments in the the sense of $UT_2$ and invoke that umbrella theory. Therefore, if relatively ill adapted alleles result from interbreeding of some sub-species populations, those populations will tend to become separate species, *ceteris paribus*. (This describes the Taubers' hypothesis of how *Chrysopa downesi* and *Chrysopa carnea* emerged as separate species from a common ancestor species. Again, the pattern of reasoning is circular.)

This may be an auspicious moment to remind ourselves of something we saw in chapter three above as we considered recursive algorithms. The mutual dependence of some umbrella theories on their sub-theories can be seen in examples such as the Taubers' sympatric speciation hypothesis. However, there is perhaps a more clear-cut way to demonstrate that a set of statements can be circular without also being empirically vacuous. We have already reviewed such a case above in the form of Wirth's solution to the eight queens problem. Recall that he found a solution to the problem by using a recursive algorithm. Or consider the following algorithm:

```
START Factorial (n)
        If n = 0
                Then Set result to 1
                Else Set result to n x Factorial (n - 1)
        Stop
END...
```

(Schneider *et al* 1978: 52)

Notice that this algorithm is anything but vacuous, even though it would be judged circular by the standards which Mills and Beatty as well as Sober employed. In the next chapter we discuss recursion as it might apply to fitness in evolutionary biology.

# Chapter Eight: A Recursive Concept of Fitness

## 1. Heuristic indicators that some "circular" concept of fitness is appropriate

One cannot read the literature of evolutionary biology without sensing circularity in the relationship between the phenomenon of selection and its alleged agent, fitness. Brandon offers a typical statement of this relation: "Selection at the level of individual organisms has as its cause differences in individual adaptedness and its effect is adap[ta]tions for individual organisms" (1978, Sober 1984c: 59). In other words, differences in adaptation (which Brandon uses as synonymous with what is here termed fitness) cause selection, but then the selection in turn becomes the cause of differentials in adaptation. This seems correct or even a matter of common sense upon a moment's reflection. If organism A is hugely better adapted (fitter) than organism B, it will not be surprising if A contributes a large number of progeny to the next generation while B produces few or no offspring. Thus the environment has selected a large part of A's genetic makeup for continuation in succeeding generations (assuming A's superior adaptedness is heritable and not an ontogenetic variation). But the relative increase of "little As" with respect to "little Bs" in the second generation will likely have an effect on the third and following generations' overall adaptedness with respect to the environment, just as adding more blue balls to a barrel than we take away red balls will increase the chance that a random draw will turn up a blue ball. In other words, adaptation causes selection in some sense, which in turn causes adaptation.

The interplay of benefit and need has been described in somewhat circular terms before (though not, so far as I know, with explicit reference to recursion). Dewey, for instance, complained of the "spin" which Victorian society had placed on the concept of evolution. To simplify, he believed that the popular reading of the late nineteenth century was to treat evolution as a progressive force driving the universe at large and humanity in particular toward a state of ultimate perfection. Although progress toward the end state is not uniform by this reading, the movement tends to be directed and linear, and the goal of the system is stasis. Dewey preferred what modern readers would consider a purer, less sentimental understanding of evolution as an ultimately directionless, unbounded and dynamic process. It is interesting to note how well one of his statements of his own position lends itself to a recursively circular interpretation.

> Adherents of the idea that betterment, growth in goodness, consists in approximation to an exhaustive, stable, immutable end or good, have been compelled to recognize the truth that in fact we envisage the good in specific terms that are relative to existing needs, and that the attainment of every specific good merges insensibly into a new condition of maladjustment with its need of a new end and a new effort. (1922: 264)

The very term "circular" as applied to the concept of fitness is a visual metaphor. As we have seen, it also seems appropriate to apply such a metaphorical term to certain biological phenomena. First there is the phenomenon of growth and reproduction itself. Again to mix Aristotelian concepts with the terminology of contemporary evolutionary biology, one can see the somewhat random creation of variation through genetic combination as an efficient cause associated with a formal cause (one of the growth stages of an organism--a typical adult, say). Aspects of the formal cause (the adult's propensity to reproduction in sexual organisms) can also be seen as an efficient cause leading to the fertilized egg of the next generation, and in turn the egg is conceived as a kind of form consisting of half of the original organism's genetic matter. These two perspectives correspond to common views taken by evolutionary biologists in looking at fitness in terms of survival, from gamete to adult capable of reproduction, and in terms of reproductive success, from adult to gamete (e.g., Sober 1984a). In short, one can see any portion of a continuous cycle as comprising efficient and material causes on the one hand (i.e., standing in a necessary causal relation to the present and subsequent states of the individual and its lineage) as well as formal and final causes on the other hand (i.e., as effects of the previous states of the organism and its ancestors). But if we reject the notion of eternal and immutable species, insisting instead on injecting the possibility or even the statistical probability of change into the process, how do we describe the progression of the cycle? (To repeat what has already been said above, it is not clear that the traditional reading of Aristotelian taxonomy is correct; cf. Lennox 1987.)

If circularity is a good metaphor for expressing a cyclical process which remains unchanging, perhaps *spiral* would be a good description for a cycle which changes very slowly. The point here is not to split verbal hairs, but rather to develop a new perspective on the process of evolution with an emphasis on fitness. It does not require a diagram to imagine how simple a model of eternal species could be: just a circle, perhaps with an arrow labeled T (for time). The diagram would indicate that past and subsequent generations have gone or will go through the exact same paths of development, with some individuals falling short of expected longevity and some exceeding it. We can conceive of a number of points at various places along the arc of the circle, with each point showing the current status of living organisms. Deceased and future organisms are not represented, nor are they of particular interest, since they are identical with presently living organisms in all essentials. What would a full arc -- "coming full circle," one might say -- represent? Of course there is no single right answer. We might choose to let the circle represent the period from birth to death, with organisms traveling the distance at various speeds.

Now if the concept of fitness is to address either survival or reproduction, any numerical reckoning of fitness must in some sense travel this circle. This is as true if we calculate fitness in terms of actual longevity and actual numbers of offspring produced as if we focus on the probability that a given organism attains a certain age or that it produces a given number of offspring. As has already been observed, the actual numbers and the probabilities are alike based on observations of real organisms' actual life cycles. But what is of even greater interest in this context is that a circle of a certain sort must be traveled even if we reject Aristotle's immutable species and seek to represent evolutionary change. Because the change is so slow, each generation of organisms will usually travel the same general cycle. Of course rapid mutations and even overnight extinction are possible, but let us assume that they are not the norm. How do we represent a "progressive cycle" such as the one which seems to be indicated here? A spiral seems to fit the bill, but perhaps we could stick with the simple diagram and simply imagine a third dimension demonstrating that there is progress, albeit a kind of progress which is manifested circularly. We will return to these metaphors presently.

## (1) An algorithmic concept of fitness

We might think of statements about fitness as though they were mathematical *functions*: the fitness of a given organism *n* would then be a *function of* its survival and/or reproductive success, or $F_n = f(S_n \text{ v } R_n)$. This equation could even stand as a definition of fitness if we were to add some sort of quantifier such as "for all n...." Unfortunately, the equation would probably prompt the accusation that it is tautologous, just as certain corresponding prose definitions did for Mills and Beatty and for Sober. But suppose that we assert the relationship between fitness on the one hand, and survival and/or reproductive success on the other, in a more dynamic fashion, referencing not just the organism *n* but also the time *t* at which we make our observation of the organism. Thus we do not presume to have a god's eye, hindsight view of the organism's total history of longevity and reproduction. Rather, we can observe an organism in the middle of its life cycle and then report what its "track record" with respect to these variables has been up to that point. The function would then look something like this:

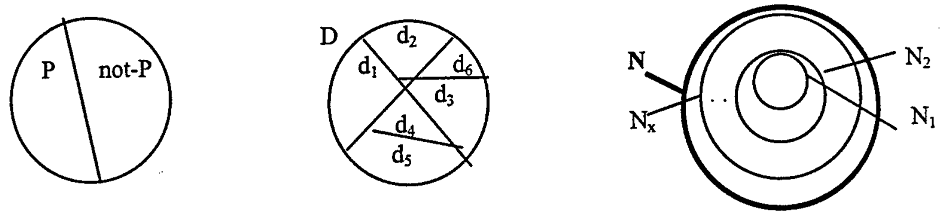
$F_{n,t} = f(S_{n,t} \text{ v } R_{n,t})$ where

$$S_{n,t} \text{ v } R_{n,t} = \sum_{x=1}^{t} S_{n,x} \text{ v } R_{n,x}$$

This format makes it easy to see that the function references a key part of itself. That is, the longevity and reproductive values are composites of their earlier values. This aspect is perhaps not so clear in common speech about fitness, but arguably the self-referential quality is nonetheless always present. When biologists observe an organism in the wild -- a tiger, say -- they may pronounce her age to be seven years and add that she has given birth to five live tiger cubs. One could say that this observation is the sum of seven annual observations, in which variables announcing age and number of offspring were updated by referring to the earlier values. (Of course it does not matter whether the observations were really made at any given frequency. What is significant here is that the phrases "seven years" and "five cubs" imply we *could* have observed the numerical progress at a particular interval.) This is what I mean by saying that fitness can be understood recursively, where "[r]ecursion means that a function can refer to or call itself. A function is said to be recursive if it refers to itself in its definition" (Frenzel 1987: 234). Recall (from chapter three) that an easy example of such a function is the computation of a factorial, where n! = n(n-1)(n-2)...(1) (ibid.: 255).

We might skip a few demonstrative steps, then, and say that a definition of fitness such as Sober's $F_s$ is really the sum of its earlier values. Each of these previously established quantities can in turn be seen as a claim about a real event, about the length of time a particular organism has lived and the number of offspring it has produced, just as n! is the abstract name of a certain function, whereas 1 or 2! are well-defined, specific values which can themselves be named n! for a given n. From this perspective $F_s$ can be an abstract definition and at the same time an empirically verifiable claim about a real event.

These speculations suggest that Sober's "surrender" in calling $F_s$ tautologous was unnecessary. As discussed above, he apparently meant that $F_s$ is analytic. But our reflections on recursive functions indicate that there is an intelligible difference between these three categories of statements -- tautologies, analytic statements, and recursive functions -- and that moreover an abstractly circular definition of fitness can also generate interesting empirical claims about fitness. A graphic depiction may help clarify the matter.



tautology
(P or not-P)

Analytic statement
(*definiendum* D $\subset$ *definiens* $d_i$)

Recursive statement
(N = f ($N_i$))

Comparison of statement "shapes"

Let us approach the same point from a different angle by accepting that $F_s$ is, as Sober suggests, a definition. Then $F_s$ can at the same time be an empirically verifiable claim once its significant semantic terms, viewed as place holders, are "filled in." In this second sense the definition of fitness is algebraic in that true and false values can be substituted for its key terms. This dual nature -- definition and claim -- emerges if we think of $F_s$ as being part of a recursive system in which new observations are continually brought into the system and tested against $F_s$ as a sort of yardstick. Thus $F_s$ tells us what it means for a given trait to be fitter than other phenotypes which could appear in its stead: it means that the trait under consideration has "a higher probability of survival and/or greater expectation of reproductive success" than other traits which could take its place in an organism's overall makeup. But what does it mean for a trait to "survive" or to have a certain "expectation of reproductive success"? There seem to be two possibilities. First, a trait might "survive" in the sense that it reappears in successive generations of an organism; and correspondingly, a trait might have a greater expectation of reproductive success for the same reason. Sober probably has in mind senses of survival and reproductive success which are tied to the survival and reproductive outlook of organisms bearing the trait. (I say "probably" because Sober's general presentation deals less with traits *per se* and more with organisms.) Throughout the following discussion I will assume that Sober intends this latter, organism-based sense when he speaks of survival and reproductive success. At the same time that $F_s$ functions as a definition, it can also serve as a template for the analysis of new data: as we observe phenomena exhibiting the longevity and reproductive record of organisms, $F_s$ serves to rank the data in a binary fashion. If a certain organism lives to x years after its birth, then $F_s$ prompts us to parse down a binary "longevity tree" until we find the right spot, that is, the spot at which the organisms represented by the branches above were longer-lived and those below shorter-lived. The combinations of traits above and below will then either give us pause and perhaps cause us to reevaluate a specific trait, or else confirm our present hypothesis. Similarly in the case of an organism's reproductive history, $F_s$ is our guideline for placing an organism in the hierarchy of its peers. Naturally these hierarchies of longevity and reproductive success are only intelligible when the environment is held "constant." (Since it is doubtful that any environment can remain truly constant, this latter condition must mean something like "within parameters agreed upon by convention.") Thus $F_s$ is not just a definition but is also an empirical claim which demands a Boolean response: either the organism exhibiting a certain trait is or is not fitter than a given sub-group of its peer within a specified environment. A graphic view of the situation would look something like Figure 1 below.

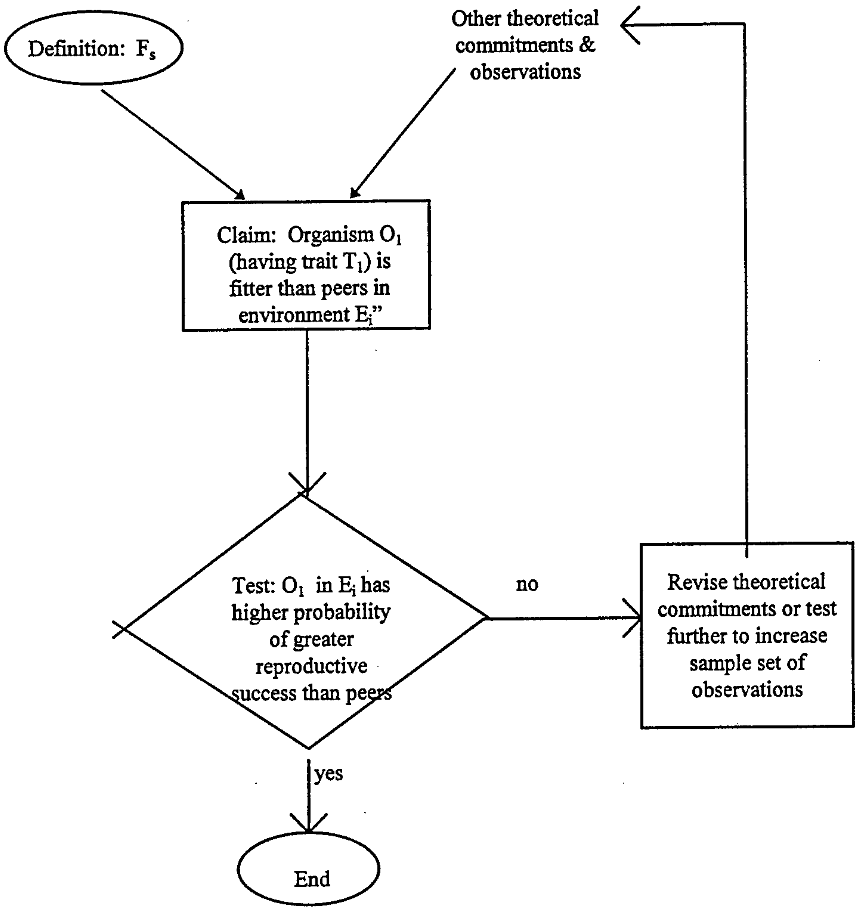

Definition: $F_s$
Other theoretical commitments & observations
Claim: Organism $O_1$ (having trait $T_1$) is fitter than peers in environment $E_i$"
Test: $O_1$ in $E_i$ has higher probability of greater reproductive success than peers
no
Revise theoretical commitments or test further to increase sample set of observations
yes
End

Figure 1: A Recursive sense of fitness referencing probability

A further complication is that the organisms in whose fitness we are interested tend to be viewed *en soi* rather than *pour soi* (to use a distinction Sartre intended in a context different from our present one) -- as things acted upon by their environment rather than as things which shape their environment as well. As has already been remarked above, the briefest reflection will tell us that in fact the causal path runs both ways.[34]

It seems to me that the diagram in Figure 1 *nearly* reflects the two ways in which the concept "fitness" is actually used by evolutionary biologists. First, fitness is employed as part of an abstract definition which refers to a "trait" as well as to "survival" and "expectation of reproductive success" in general. In the case of this diagram, the definition used is simply Sober's $F_s$. But the "shell" of the definition can be "filled in" to check the nature of specific traits exhibited by actual organisms in given environments. The "truth value" of fitness as an attribute of a given trait in this second usage is liable to constant amendment, while "fitness" as it stands in a definition such as $F_s$ is not. In this sense, we take $F_s$ as an hypothesis and continually revise it until the test (represented by the diamond in the flow chart) has a positive outcome. The "negative" path between this test and the box showing our commitments and observations may be traveled over and over again as we revise our views of the situation. This updating of specific information which is "fed" through a loop whose form remains constant prompts the term "recursive" to describe how the process works.

But I believe that the diagram above is flawed in at least one sense: it could be simpler. For suppose that we omitted any reference to probability. Presumably the flow chart would nonetheless be accurate, or even more accurate than it now is. In other words, although $F_s$ *qua definition* makes reference to probability, the *claim* that a particular organism is fitter than another in a given environment *need* not mention probability and perhaps *should* not do so. Without such mention, the diagram -- and the conception it represents -- is more parsimonious. The role which probability (propensity) was to play is taken over by recursive repetition. If an organism which we deem to be fitter than another because of its traits and our past experience turns out not to be so with respect to *real* longevity or in terms of *actual* offspring produced, then we modify one of our "background" commitments. We might choose to revise our appraisal of the traits in question, or we might decide to say that chance caused the anomaly. But the *process* of reasoning depicted in Figure 1 still works as before even though the reference to probability is omitted (Figure 2).

But what of the difficulties which Mills and Beatty (1979) and Beatty (1980), among others, attribute to a concept of fitness defined as *actual* longevity or *actual* reproductive success? We will return to this issue below, but for now suffice it to say that determining what constitutes a "propensity" is (as we have seen) extremely problematic. Using the "expected number" of offspring (defined as the average of the weighted values of possible numbers of offspring) is perhaps the best response to these difficulties, but it is far from adequate. The expected number of offspring can change as new data become known, yet propensity as a property inherent in an organism (as Mills and Beatty assert it to be) should not be subject to change based on new data unless he organism or its environment changes. Only our estimate of that value should be variable. For this reason it seems desirable to remove reference to probability from a definition of fitness as propensity (Figure 2), and instead include the probabilistic aspect of the reckoning elsewhere in the process of analysis.

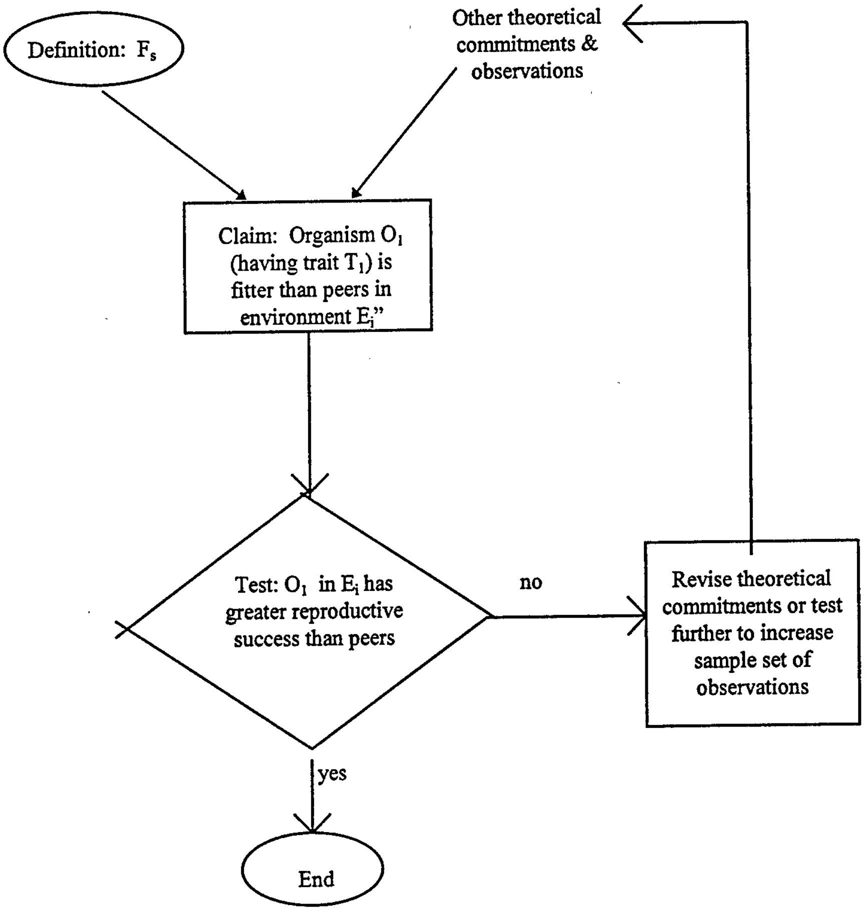

Definition: $F_s$
Other theoretical commitments & observations
Claim: Organism $O_1$ (having trait $T_1$) is fitter than peers in environment $E_i$"
Test: $O_1$ in $E_i$ has greater reproductive success than peers
no
Revise theoretical commitments or test further to increase sample set of observations
yes
End

Figure 2: A Recursive sense of fitness which does not reference probability

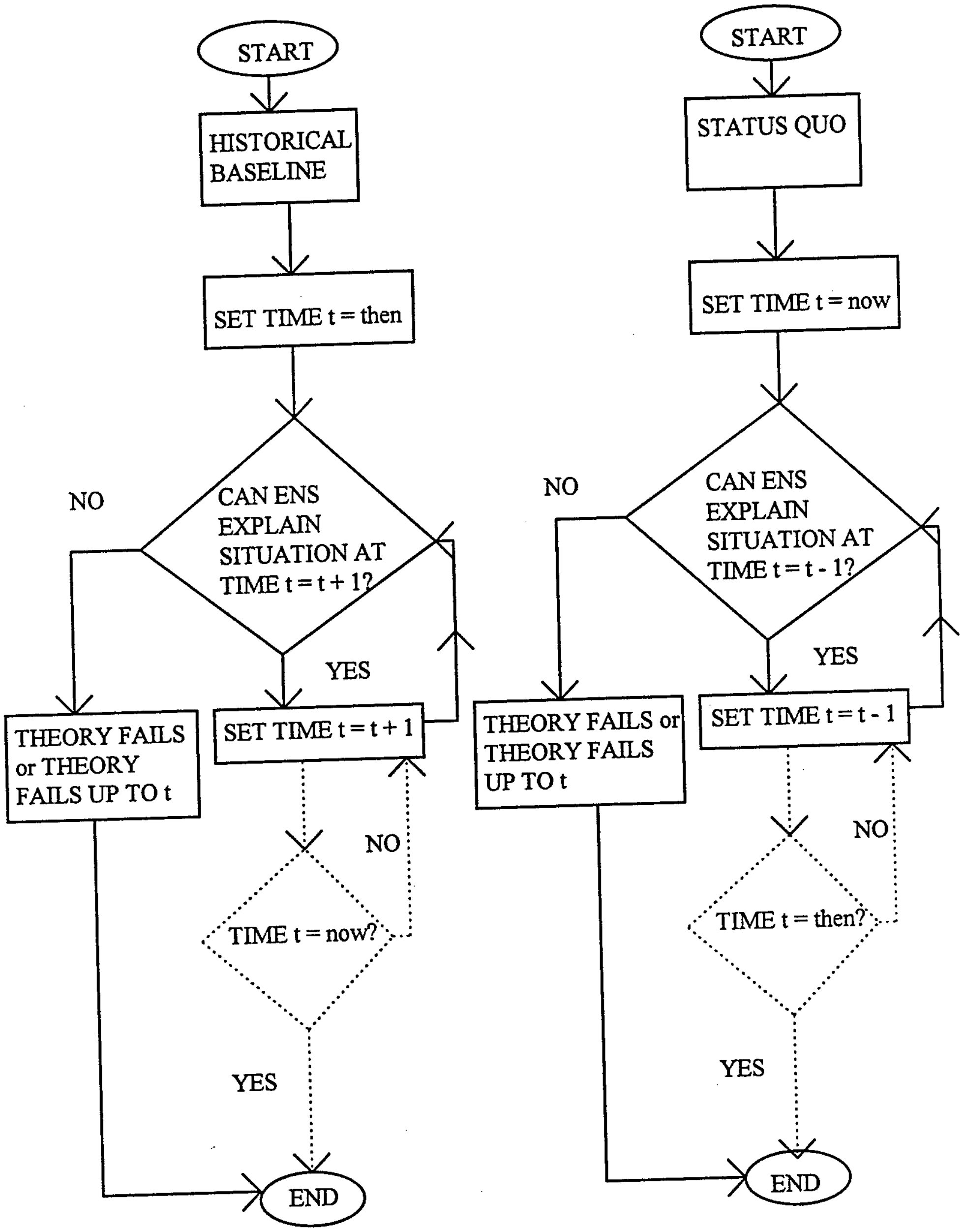


Figure 3: A recursive model for employing Evolution by Natural Selection (ENS)

Notes on the diagrams:

1. The diagram on the left can be seen as a way of working forward from the present of from a point in the past. The model depicted can end at any time, but apparently it *must* end if it fails to explain the transition which takes place between any two "steps." For instance, if we find that evolution by natural selection (ENS) cannot explain an instance where two distinct species having a common ancestor would be equally well adapted in a certain environment, then presumably we would have to appeal to some other model to explain the difference in

structure. The one- and two-horned rhinosceri cited by Lewontin (1978) would be an example.

2. The diagram on the right works in the opposite direction -- from the present toward a point in the past.

3. With slight modification the pictures might also be taken to represent a model of ENS based on mathematical induction:

Given an ordered set of "steps" 1...x,

a. Assume ENS is true for some phenomenon. That is, there is an empirically verifiable case of evolution or of possible evolution by natural selection.

b. If ENS governs the genetic development of a species at time t = n, then ENS governs the further genetic development of the same species at time t = n + 1.

c. Then ENS is a possible explanation of existing species and of ongoing changes within these species. Furthermore, ENS viewed in this way is neither tautological nor analytic, despite the fact that its truth is assumed in the course of step 2.

Perhaps we can go even further than Figure 2. Suppose that we try to construct a diagram representing single observations of longevity or reproduction. That is, imagine that we want a schematic showing how we appraise a single organism in terms of its longevity or reproductive success. In that case it would not be merely unnecessary to include a reference to probability in the "decision diamond"; it would be inaccurate to do so, since the single observation alone cannot indicate a general probability reflecting on the longevity or reproductive success of the organisms in the taxon to which the individual belongs. A judgment as to the probability of the organism's relative longevity or reproductive success emerges from *iterations* of the test itself. But each individual test can function adequately without a claim about the probability of a certain longevity or degree of success in reproduction.

This may raise suspicion as to whether the definition of fitness, $F_s$, must refer to probability. The criterion of parsimony would again argue that the answer is no: $F_s$ can make such a reference, but a workable concept of fitness does not require it. Graphically, a simpler conception of fitness would look like Figure 2 or 4.

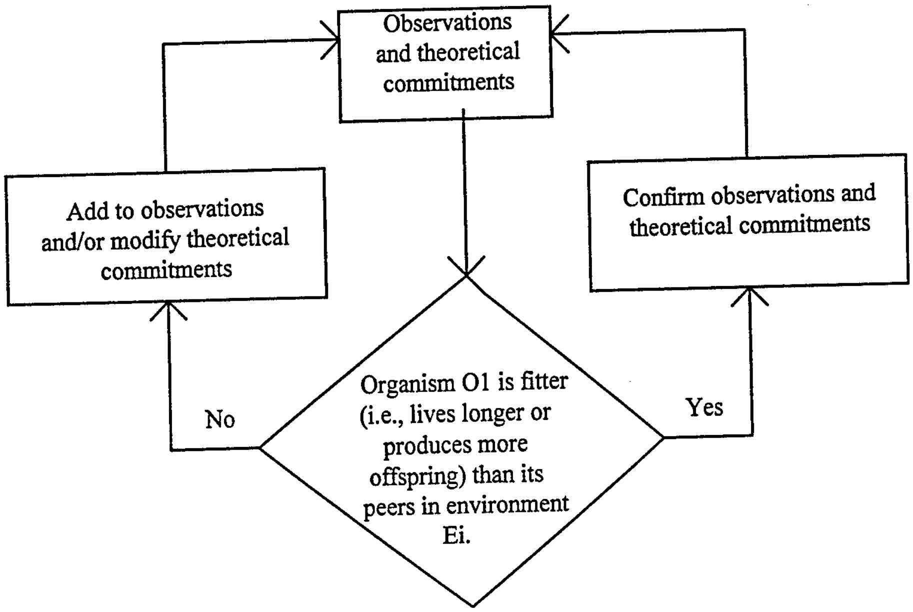

Observations and theoretical commitments
Add to observations and/or modify theoretical commitments
Confirm observations and theoretical commitments
Organism O1 is fitter (i.e., lives longer or produces more offspring) than its peers in environment Ei.
No
Yes

Figure 4: Recursive test of fitness.

The process depicted in these figures still includes a probabilistic element, but any reference to probability is provided by recursive iteration or included in the theoretical commitments (depending on perspective). The notion of fitness can function adequately in both its definition and in a claim about an individual organism in a certain environment without referencing probability. It also seems appropriate that the process in Figure 2 need not hit an exit condition and that the process in Figure 4 does not include specific start and end points. This is because the process of inquiring about the fitness of specific organisms in given environments can go on indefinitely, with each observation potentially modifying or enlarging the basic stock of observational and theoretical knowledge. Modification of the definition of fitness is not an option here.

It may be objected that the *definition* of fitness cannot do without some reference to probability -- that is, to a *tendency* or *propensity* to longevity and reproductive success -- lest the old bugaboo, chance, make theorists' lives miserable once again. What we want to avoid is being driven to assert a paradox by a definition of fitness which is tied too closely to what actually happens. As we have seen, that unhappy circumstance obtains in situations where, for instance, one of a pair of identical twins is killed by lightning before having had the chance to reproduce (Mills and Beatty 1979; Beatty 1980). If our definition of fitness is tied to actual survival and reproductive success, it has been argued, we would have to dub the surviving twin fitter than its identical sibling. But note that Sober's $F_s$ refers to the survival and reproductive expectations of *traits*, not of individuals. The fact that an individual bearing a trait T expired at a tender age and without leaving descendants is a reality affecting the individual bearing a trait; the trait itself must be viewed as independent of its bearer. Since a trait does not itself "live" or "reproduce" independently of the organisms which bear it, it is in some sense immune to chance. We do not have to limit ourselves to using the language of probability in the case of traits -- we do not have to say that a certain trait tends (or "has a propensity") to be fitter than another trait. Instead, we are justified in saying that one trait *is* fitter than another, given the data available up to a certain point.

Speaking in terms of propensities and tendencies is not necessary in the case of traits, and therefore it is not necessary in the case of the recursive models of fitness depicted above. But is this simply an academic point, or is there anything to be gained by removing reference to probability from the definition $F_s$ and from the recursive model? As indicated already, the chief gain is in simplicity. It seems desirable to picture the investigative process which evolutionary biologists actually use when measuring fitness in as economical a fashion as possible. But there is another reason as well, one which is related to what will be said below about Wachbroit's concept of "biological normality." The point to be made in that discussion is that biological normality is a derivative of statistical normality, not a separate sort of normality (a point reminiscent of the insistence that fitness be ascribed to groups of organisms rather than individuals; q.v. Ettinger *et al* 1990). In a similar sense, a concept of fitness which references probability seems to be ultimately an extrapolation of a more basic sense of the term, one which is based on organism's actual longevity or the actual number of offspring it produces. Before we can predict from traits the probability of a certain longevity or level of reproductive success, it is apparently necessary to see which organisms actually live longest and which actually produce the most offspring. For that reason it seems to me desirable to keep reference to probability -- which I admit must come into play somewhere in the process of analysis -- associated with general theoretical commitments and not specifically with fitness.

There is yet another reason why we should think of the use of fitness in evolutionary biology in the recursive sense as depicted above, where probability does not figure in the definition of fitness itself. This reason has to do with the question of what information is relevant to quantifying probabilities. I believe it is the case that our intuitive sense of the connection between probability and fitness, or propensity and fitness (which amounts to the same thing in this case), can mislead us. We have already seen examples of this above -- in interpreting Copi's questions with regard to two-card hands, for instance, and in the examples involving dogs and penny-piglets. There are likewise other aspects of life relevant to evolution which we are better off evaluating recursively without reference to rather than in terms of propensity. One such is cooperation.

## (2) Altruism and the Parole Board's Conundrum

The "altruism problem" is a theory-laden bother; *per se* it isn't irksome at all. Generations of animal-watchers have accepted that organisms sometimes sacrifice themselves for the sake of other, usually related organisms. Some organisms have even seemed to non- or semi-scientific observers to offer their own resources to further the survival of other species. Thoreau, for instance, mentions "cowbirds and cuckoos, which lay their eggs in nests which other birds have built" (Thoreau 1854: 36); the other birds, in turn, would hatch the eggs. We start caring about apparent altruism when we insist as a theoretical commitment that a foundational selfishness must be apparent in all our observations. At that point something has to give. We must surrender selfishness or else altruism, or find a theory which can accommodate both together. For Dawkins (1976, 1986, 1989, 1995; cf. also Glance and Huberman 1994), the selfishness remains and the observations themselves are called into question. The concept of selfishness is tied-up with a self, in this case conceived as a gladiator in the survival and reproduction arena. What if individuals were not the agents striving for hegemony on the survival and reproductive fronts? What if they were merely the puppets whose strings are pulled by the real agents? Dawkins' puppet masters are genes. In his theory the "level" of agency, to use a spatial metaphor, has changed from individual or group to something much narrower, namely, a constituent part of the individual. The intermittent altruism of individuals is thus a façade belying the consistent selfishness of genes; when individuals appear to act selfishly, on the other hand, their deeds are a true reflection of their genes' motives. ("Motivation" must be taken metaphorically here.)

But apart from a specific theory such as Dawkins', how do we evaluate the fitness of organisms when they are involved in cooperative contexts? Do we appeal to probability, or is there a better way? Let us approach this question by looking more closely at altruism itself.

Altruism can arguably exist in two forms. First, there may be a pure selflessness, devoid of any ulterior motives, which is pursued as an end in itself. But altruism can also exist as a means by which self-interest is pursued. By the phrase "enlightened self-interest" we often describe such an essentially selfish kind of apparent selflessness, in which the agent sacrifices immediate gratification for a greater gain in the long term.[35] It is difficult to say what can count as "gain" in this second sense. Suppose that a bear sacrifices herself in order to protect her cubs. An observer might explain the phenomenon by claiming that the bear acted from pure altruism. Alternatively, it might be claimed that the bear had a selfish interest in protecting her cubs based on some goal relevant at a level of existence other than that of the individual -- for instance, group survival or perpetuation of genes through the "vehicle" of offspring. Paradoxical though it may sound to claim an agent would pursue its "self"-interest to the point of ending its own existence, at least the claim is not clearly false if we place the beneficiary on a wholly different level than the self. Self-interest seems to require that the agent, or at least some object connected with the agent (on another level), exist to enjoy the pay-off of the course of action pursued. That is, if agent A pursues a course of action which will eventuate in a goal G, then apparently the pursuit can be selfish in just two senses. First, it is possible that the pursuit itself is gratifying to the agent regardless of the outcome. Alternatively, the goal can be the object of desire. (In its form this distinction is at least as old as *Republic* II, 357b ff.) But if achieving the goal is the pay-off, then presumably the agent undertakes the action believing it may survive. In either case -- where the action is gratifying in itself and where the agent does something apparently selfless, even risking its life, on the chance that it will enjoy a gratifying outcome -- it appears that "selfish altruism" is possible. But sometimes the selfish action seems to be willingly undertaken even if instinct is the primary motive. For instance, although a she-bear may defend her cubs instinctively and therefore without the need of prior, conscious reflection, *she might have done otherwise*. Sometimes she may retreat. Her actions are not wholly "hard-wired."

Darwinian evolution depends on self-interest, though perhaps not of a conscious sort. By this theory, the environment places pressure on the individual or the group or perhaps on some other "level" of the spectrum of life (e.g., the gene). Organisms respond by changing their morphology or their behavior. But this evolutionary change is arguably *unconscious* -- that is, unintentional -- even if its manifestations are purposeful acts. (I say "arguably" since controlled breeding through social mores as well as future manipulations of the human genome might be viewed as evolutionary change and grounded either in unintentional genetic-based motives or in an intentional desire to change the course of the species' development.) Perhaps among the ancestors of bears there were timid creatures who, during one period in the course of their evolution, would have abandoned their young immediately rather than run the least risk of harm. Later these creatures evolved into fierce defenders of their young, but the evolution itself was not intentional. Rather, the evolutionary change in behavior occurred as unintentionally as any morphological or behavioral change -- from short, scaly fins to long legs tipped with paws or hooves, say. By a Darwinian account, it was never the case that an organism said to itself something in the nature of, "Hey, these flippers won't do for ground transportation. I better develop something a lot longer and a bit more durable where it touches the ground." The case becomes more complicated when we view the kinds of behavior describable as selfish or altruistic as being the *direct* result not of instinct but of reflection. (We leave aside the question of whether reflection may be *indirectly* linked to instinct.)

Tucker's well-known prisoner's dilemma (q.v. Davis 1970: 93- 103; Martin 1992: 175) considers cooperation from the perspective of prisoners who must decide whether to cooperate with one another or not. Obviously there are reasons why they should cooperate, but there are also selfish motives which urge an independent course of action (Hume 1740: 485; Árdal 1989: 175). The alternatives in Tucker's version can be schematized as follows (cf. ibid.: 94):

| suspect 1<br>suspect 2 | confesses | does not confess |
|---|---|---|
| confesses | both serve 5 yrs | serves 20<br>goes free |
| does not confess | goes free<br>serves 20 | both serve 1 yr |

From a defendant's viewpoint in Tucker's conception, there are three key questions. One of these questions is obviously selfish: What will I get out of cooperation with my fellow prisoner as compared with pursuing one of the other alternatives? The second question may or may not be purely altruistic, because even though it takes a communal perspective, that apparent concern for one's fellow captive may be altruistic only in a short-term, ultimately self-interested sense. That question is, What do *we* get out of the various alternative courses of action, including cooperation? If this question is indeed ultimately selfish, then it is clear why one would ask it as well as the first question. Finally there is the question, Can I trust the other guy? Or perhaps talk of trust is premature or even unnecessary. What the prisoner really needs to know is how the other guy thinks and what he will therefore do? Each prisoner must know or guess what the other is thinking -- how the other will evaluate the alternative courses of action -- in order to decide rationally which course of action he himself should take. Without such knowledge, the decision to "sing" or not becomes more like a flip of the coin.

But the prisoner's dilemma incorporates another epistemological aspect besides the one having to do with what the prisoners know about each other's possible actions. There is also the question of what the observer of the dilemma knows and how she knows it. Consider the "forces" at work in the prisoner's dilemma. The one which normally engages our attention is the motivation of each prisoner. Presumably that force can be dubbed "self-interest," but such identification does not automatically clarify what the prisoners are going to do. However, this motivation of self-interest makes sense only against a backdrop of regularly applied penalties. This penal code functions as a kind of environment, one which reacts to a each action on the prisoners' parts with a specific penalty. The penalty for a given action is fixed and can be expressed as part of a triplet {a1, a2, p} in which the complementary elements are the respective actions of each of the prisoners. From the point of view of a single prisoner, the action of the other prisoner could be conceived of as part of the environment, so that the combination of action and environment yields the penalty: {a, E, p}.

A model of Darwinian evolution similarly links an organism's specific structure and behavior with a certain outcome in a given stable environment. Corresponding to {(prisoner 1 does not confess), (prisoner 2 does not confess and rules as specified above), (both prisoners jailed for one year)} in the prisoner's dilemma, one would find something like {(deer x is slow), (environment with many fast predators), (deer x is short lived)} in Darwinian evolution. It should also be noted that although the possibility of a positive outcome being associated with an organismic phenomenon-environment pair is more obvious in the case of Darwinian evolution, whether an event is viewed as positive or negative depends largely on the observer's perspective in the case of either model. Thus *fast deer : long-lived deer :: cooperating prisoner : prisoner jailed for one year* sounds essentially positive because of the way the relationships are expressed; but the deer still dies in spite of its speed and the prisoner still does hard time despite making the right choice (or a lucky guess) vis-à-vis his comrade.

In one aspect, however, the relationship between the two models is unclear. In the case of the prisoner's dilemma, there is no question as to how the prisoners and the observer know the penalty associated with each possible course of action; indeed, the scenario takes its character as a certain kind of dilemma precisely because one prisoner must guess how another prisoner will act given the same knowledge of the penalties associated with each option. The dilemma would be quite different if one prisoner was not sure whether the other knew the rules, or knew the same set of rules, before making a decision. But in the case of an analogous model of Darwinian evolution, the prisoner-analogs -- namely, organisms -- presumably need not have any conscious knowledge of the "rules of the game" at all. Except in the case of some humans, an individual organism may live and die entirely without knowledge that a certain genetic makeup dooms her to sterility and an abnormally short life span. The observer's knowledge is likewise questionable. How is it that we come to believe that certain rules have applied to evolutionary phenomena?

We will ignore the first issue -- whether and how any organisms may understand the way environmental "rules" affect themselves, personally -- and focus on the second question, that of how an observer draws conclusions about the rules governing the evolution of organisms. Consistent with the scheme sketched above, this question can be translated into one about triplets of factors: How do we construct a triplet such as {(organismic structure and/or behavior), (stable environment), (probable outcome)}?

Interesting and relevant though the standard form of the prisoner's dilemma is (cf. Glance and Huberman 1994), in trying to answer this question we may profit by adopting a different perspective -- that of a parole board which must try to discern the motives of the prisoners after the fact, no matter what they decide. Such a role more closely approximates that of the community of evolutionary biologists as they try to appraise how the social behavior of organisms either does or does not advance the causes of individuals (or the interests, metaphorically speaking, of a group or a collection of selfish genes).

Let us stipulate that in the parole board's conundrum, the rules are similar to those of the prisoner's dilemma, except in this case jail time is already being served by one or both of the initial suspects, depending on what they chose to do. The decision facing the parole board is not how best to serve its own self-interest (except in a very remote way, that is, in so far as the board and its members thrive whenever organs of the public welfare do their jobs well). Rather, the board must try to discern the character of each prisoner who has come up for parole (as the defendant or defendants jailed by Tucker's original problem surely will at some point). As in the real-life scenarios played-out in many Western prisons, inmates' sentences are usually only upper limits; most will go free before they have served the amount of time which a judge has set as the maximum period of incarceration. It then becomes a parole board's prerogative (doubtless guided by legislative and judicial decisions, the current degree of prison overcrowding, as well as the social climate overall) to shorten or maintain sentences depending on aggravating or mitigating circumstances in a given prisoner's personal history. For instance, if a parole board finds evidence that a defendant has demonstrated remorse and a cooperative attitude -- say, by confessing -- then it will be more inclined to commute a prison term. But if the parole board finds the prisoner uncooperative and remorseless -- the conclusion it may draw if the defendant remained silent but was nonetheless convicted -- then it will tend to maintain the maximum sentence. Were the scenario in Tucker's dilemma to unfold, say, in the U.S., some prisoners would serve their entire twenty-, five-, or one-year sentences as surely as Charles Manson will serve his natural life (that is, barring miracles), while others would go free much sooner. It all depends on how motives are appraised by observers who do not have direct access to mental processes but instead must infer them from present behavior and past experience. Presumably the same kind of inferential process would need to be carried out if the prisoners had no real motives but instead acted from grounds we might term instinctual or pre-programmed. Since no observer has access to the intentional state of the prisoners, it does not really matter whether that state is rational or instinctual (at least not for the purposes of analysis *per se*, though of course a penal philosophy would likely treat rational agents differently than non-rational ones) so long as the observer can discern some sort of regularity. Finding a pattern of action may be easier given one hypothesis (e.g., instinct) rather than another (e.g., rational decision-making), but in the end the observer is always guessing. Most importantly, even though there is no way of penetrating to the intentional state of the prisoners apart from inference, that does not mean that we cannot say something about the way the prisoners are motivated. (This is one point which must be granted to Dennett, I think, regardless of how his argument as a whole is appraised. Cf. 1991: 441 - 8).

The point of offering the parole board's conundrum is simply to suggest that a model of altruism versus selfishness, or indeed of any aspects of behavior, can be built recursively perhaps more easily than any other way. Consider the case of an overworked parole board which attempts to clear its backlog of cases and conform to a tight personnel budget by building an automated system. The goal of this essentially behaviorist system would simply be to shorten the amount of time necessary to review all the relevant parameters in a prisoner's history. The board may have despaired of effectively interviewing parole candidates, having long ago concluded that most will say anything in order to win early release. Thus the board's philosophy is simple. "Past actions are what counts!" the chairman of the board has cautioned the other members as well as the team of computer specialists responsible for implementing the new system. How, then, will the knowledge base of such an expert system be compiled? Almost certainly the key to the algorithm will be to consider each prisoner as a "function" of her own past "selves": $\mathit{ParoleCandidate}_{\mathrm{time=t}} = \mathrm{f}\,(\mathit{ParoleCandidate}_{\mathrm{time=t-i}})$, $\mathrm{i} \geq 1$ , where the particular function may relate the various moments of a prisoner's life to one another in virtually any way the board wants it to. The key point is that the progression of the prisoner's life can be naturally analyzed in a way that makes every successive moment or episode a function of the previous moment. Once again it should be stressed that the algorithm *could* be conceived and implemented in a non-recursive way. Nonetheless, a recursive means of building the system's knowledge base seems not only natural -- as it would arguably have been at the end of any previous century and not just at the end of our own -- but currently practical, too. To repeat, the board's appraisal of a parole candidate's fitness to be freed is analogous to the job of appraising the fitness of organisms in general.

## 2. The similarity criterion

Let us assume for a moment that Wirth's definition (discussed in a previous chapter) uses the phrase "defined in terms of itself" as a shorthand way of saying that the recursive object is defined in terms of something *similar* to itself. This raises obvious questions: What constitutes "similar"? What criteria do we use to deem the *definiens* sufficiently similar to the *definiendum* to fulfill Wirth's definition? Because we think of the recursive object as being defined by a serial (temporal) progression of identical or similar objects, it makes sense to look at the object as being dynamic -- what we might call a process or method rather than a thing which remains constant across time.

### (1) Recursion as object and method

It has already been mentioned that the chapter in which Wirth presents his definition of recursive object is entitled "Recursive Algorithms." This is noteworthy because although an algorithm can be thought of as an object, it is not like rectangles, Russian dolls, and advertisements in the pages of magazines (though television commercials may be a different matter). We think of an algorithm as being dynamic or directive. One sense of the word evokes the image of something that plays itself out over a period of time as it *does* something. "I ran the data through a new algorithm this morning," says one researcher to another. A second sense of the term is roughly synonymous with "recipe," only the range of uses transcends the culinary: "Use the right algorithm and the job will be a lot easier." The nature of the job does not really matter.

Extrapolating from Wirth's definition of recursive *object*, it appears that a recursive *algorithm* would be one which "partially consists of or is defined in terms of itself." Here we must guard against a problem analogous to the one we encountered above, where the mere repetition of a pattern is sometimes (mistakenly) called recursion. In employing recursive *method*, on the other hand, it is the *activity* of selecting a proper subset of qualities from a constant object which gives meaning to the phrase "partially consisting of itself." Wirth's key definition straddles the difference between object and method. On the one hand, recursion as self-reference deals not with being (consisting) but rather with doing or acting (referring). In the same way, partially consisting recursion *qua* method can be seen as the *action* of serially abstracting various subsets of an object's overall being. On the other hand, the self-referring function of recursion is centered in the recursive object itself; it is not a facet of an observer's method.

### (2) Orders of recursion

Recall Ridley's (1985) example illustrating how sympatric speciation (speciation without geographic separation) can occur. The details have already been discussed at greater length above, so a bare outline will suffice for present purposes. The environment in question is a single locale in the sense that its borders on a map are fixed. But within these borders lie more than one habitat suitable for a certain kind of insect. To simplify, let us call these the dark habitat, abbreviated D, and the light habitat, L. D consists of coniferous trees with dark needles and other areas of the temperate forest which form a dark background. Certain kinds of moss might qualify, along with the bark of many trees and perhaps a few dark-leaved deciduous bushes as well. L, the light portion of the environment, is more or less the complement of D:

whatever is of a light color, especially pale green, constitutes L. Grasses, light colored leaves, pale mosses -- all belong to L. As for the insects in question, let us simplify a bit by identifying two species, $S_L$ and $S_D$, which are very nearly identical except that one is of a light color and the other dark. The predators of $S_L$ and $S_D$ identify their victims visually, so as we would expect, $S_L$ lives almost exclusively in habitat L while $S_D$ limits itself to D.

Thus the phenomenological stage is set, and now comes the argument for sympatric speciation. Let us suppose that there was a parent species P, now probably extinct, which inhabited the geographical locale. Members of P could look like $S_L$ or $S_D$ or something in between, and the species' range was the whole of the environment, that is, both "color habitats." We assume that P possessed "color alleles" permitting, say, three possible phenotypes: dark, light, and in between. Assuming further that the insects are able to identify the principal color of their immediate surroundings and the color of potential mates, and supposing the heritability of corresponding behavioral phenotypes codifiable as simple rules -- "stay on or near dark (or light) surfaces as much as possible" and "mate with organisms of your own color" -- we have the elements of long-term differential survival and reproduction. The end of the story is predictable: eventually the phenotypes for an in-between color disappear as breeding among members of the P species becomes non-random with respect to color. What we can call the degree of randomness decreases until the breeding populations are rigorously demarcated along color lines. Or at least that is one scenario explaining the existence of $S_L$ and $S_D$. Even though members of the two groups share the same environment and can interbreed under laboratory conditions, they are considered to represent two distinct species.

The pattern of reasoning driving this model of sympatric speciation can be expressed negatively or positively. The negative version sketches the destruction of phenotypes which make a certain kind of organism more vulnerable to predation than competing phenotypes in the same environment. The positive side of the same coin focuses on the rise and eventual dominance of advantageous phenotypes. (Here "environment" is taken in the most restrictive sense applicable. In the example above, for instance, we call the geographical area the "environment," else we would not have an example of sympatric speciation. But in both models of sympatric speciation we must treat the "environment" as synonymous with the source of selective forces. In the negative version, of course, the total environment comes into play -- for instance, when a dark green insect remains too long on a light green leaf and draws the attention of a passing bird. But in the positive model we stress the selective beneficence of the environment more narrowly defined for organisms which are well adapted to it.)

It is easy to see either model as amounting to a recursive test of the compatibility of single alleles with environmental conditions. Naturally we have to assume a *ceteris paribus* clause is in effect so that we can isolate individual phenotypes while holding the environment and other aspects of the organism constant. The test of a given allele is really a test of the organism which is defined as the aggregate of its alleles: Is the organism with that allele better suited to its environment than with a competitor allele? Of course there is no absolute answer to this wording of what is apparently the key question. Two factors complicate the issue. First, the organism may be less adapted with a given allele and any corresponding phenotype than without, *given the rest of the organism's configuration.* But if we change that configuration, then the allele in question may be more beneficial to the organism than any of its competitor alleles would be. Secondly, the answer must take into account that the interaction between organism and selective forces is stochastic rather than deterministic. What we have to reckon with, then, is recursion embedded in recursion. As Mills and Beatty have long since pointed out (1979), statistical propensity to enjoy a certain longevity and degree of reproductive success in a given environment is one way or interpreting fitness.

But in fact fitness is not so one-dimensional. We might think of separate algorithms in an embedded hierarchy running concurrently. The innermost algorithm -- easy to conceive as functioning recursively -- simply permutes alleles through some mechanism such as recombination. A parallel level does the same for environmental conditions: all the different combinations of environmental conditions are catalogued. The next level pairs genotypic permutations with environmental conditions. A further level matches the resulting pairs with real reproductive rates to create yet more pairs. Where is fitness to be found in such a process? Before answering the question, it may help to observe that binary expressions themselves are sometimes considered to exhibit what Wirth calls a "nested, recursive structure." He pictures x + y and x - (y * z), for instance, with following, self-explanatory Boolean diagrams:

<table><tr><td colspan="2">+</td></tr><tr><td>T</td><td>x</td></tr><tr><td>T</td><td>y</td></tr></table>

<table><tr><td colspan="3">-</td></tr><tr><td>T</td><td colspan="2">x</td></tr><tr><td rowspan="3">F</td><td colspan="2">*</td></tr><tr><td>T</td><td>y</td></tr><tr><td>T</td><td>z</td></tr></table>

(Wirth 1986: 172 - 173)

The recursive character emerges not from the picture *per se*, but rather from the procedure which would be used to evaluate such an object. In the case of x + y, a procedure might first note that the operation is binary. Thus at some point it must be true that a Boolean condition will be met twice. That is the case on one single level in this example: the first and second tests both yield an affirmative, and with that, the recursive algorithm reaches its end and the recursive object is exhausted. In the case of x - (y * z), the algorithm must proceed to a second level. One can construct or imagine examples of immense complexity which use the same cyclic process of evaluation. First an operator is found and evaluated for type. If the current level provides sufficient operands, the operation is carried out. If not, the algorithm begins again at the next level down.

Let us try to see how a similarly recursive procedure might proceed in a context involving fitness. For simplicity's sake we can use a hypothetical machine as we attempt to visualize a model of embedded recursion. Suppose a certain "level" of a device has just two qualities (analogous to genotypes) which can differ from one device to the next, and each of those two qualities has a range of just two possible states (alleles). A sail boat, let us say, has a single mast. The mast either does or does not have a sail. Similarly, the boat either does or does not have a bilge pump. To repeat, those are the only two aspects which can vary from boat to boat. Let us imagine further that weather can be evaluated by two all-or-nothing dichotomies: wind vs. calm and rain vs. fair. Suppose we construct every conceivable tetrad of these "genotypes" by permuting the possible "alleles." (Generally one does not apply such terms to environmental conditions, but we can allow ourselves to do so for the sake of notational economy.) We will find 16 ($2^4$) foursomes, e.g., sail-and-pump-

and-windy-and-raining, sail-and-pump-and-calm-and-raining. Following the pattern of Wirth's example, the following schematization expresses one foursome:

<table>
<tr><td colspan="4">and</td></tr>
<tr><td>T</td><td colspan="3">SAIL</td></tr>
<tr><td rowspan="5">F</td><td colspan="3">and</td></tr>
<tr><td>T</td><td colspan="2">PUMP</td></tr>
<tr><td rowspan="3">F</td><td colspan="2">and</td></tr>
<tr><td>T</td><td>RAIN</td></tr>
<tr><td>T</td><td>WIND</td></tr>
</table>

Here "and" is treated as a binary operator (as - and * are in the case of Wirth's diagram for the expression x - (y * z)).

There are two ways of "simplifying" the schematic. One is to increase the "order" of the operators. If we were to treat the ampersand sign, "&", not as a binary operator but rather as one taking any number of arguments (operands), then for the tetrad sail-&-pump-&-rain-&-wind we could construct this schematic:

<table>
<tr><td colspan="2">&</td></tr>
<tr><td>T</td><td>SAIL</td></tr>
<tr><td>T</td><td>PUMP</td></tr>
<tr><td>T</td><td>RAIN</td></tr>
<tr><td>T</td><td>WIND</td></tr>
</table>

The other means of simplifying is to limit the semantic range by introducing logical operators, so that the vocabulary need not contain "neutral" terms for, say, rain and calm; rather, there is simply rain or not-rain in our example. For students of evolutionary phenomena in general (not just for evolutionary biologists), such means are helpful. Chomsky and Halle (1968), for instance, schematize the evolution of English phonetics using three vowel heights -- low, mid, and high -- and then negating those which do not apply. They similarly negate other qualities such as tenseness (the negation meaning "lax"). Thus the diphthongization of the vowel in the word "town" can be schematized by a diagram involving the terms "- cons", "- voc" and "+ tense", among others, to indicate that the diphthong is a tense sound, but it is neither a consonant nor a pure vowel sound (Chomsky and Halle 1968: 274).

Negation is just one logical means of limiting a system's vocabulary. Any number of other modifiers might be employed, though it is not our business at present to say exactly how many. Presumably we could use terms derivable from the ten or so Aristotelian categories to modify existing morphemes in a vocabulary, but of course the precision of such usages is always an issue. The story goes that an American academic, invited to lecture in Germany for the first time, was introduced to the members of a university faculty. After shaking hands with a line of Herr Doktor Xs and Frau Professor Doktor Ys, the American sought clarification: Did the title of address "Dr." indicate a person without tenure, such as the assistant profs at many American universities, while "Prof." applied to those with a permanent position, corresponding to American associate and full professors? No, he was told, a professor may not have a "Stelle" at all. The title is earned through academic work beyond the doctorate. It is rather as though someone had earned two doctorates, his host summarized. The explanation sufficed until Herr Dr. Dr. Z was introduced.

One result of this discussion is to show the difficulty of localizing a nexus of transition and therefore of determining where and how a reduction -- if any is possible -- can take place. That does not go beyond the conclusions of chapter two, but it is important that we bear in mind the odd character of reduction in recursive contexts. The primary point here is that if we conceive organisms (e.g., diploid ones) as conglomerations of pairs of genes or of present versus absent phenotypes, we can represent those conceptions rather easily as recursive objects.

### (3) The Cinderella level: fits and fitnesses

We all know the tale, but it contains a moral for students of fitness which justifies brief repetition. There are many versions; this one will do as well as any. *Deus ex machina* brings a poor maid, Cinderella, to the local prince's masque ball. In the course of the evening the two fall in love. But for Cinderalla the cost of admission was a rigid curfew: at midnight the sumptuous trappings will degenerate to their former states -- coach back to pumpkin, horses to mice, coachmen to rats and lizards. Even Cinderella's dress will go from richness back to rags on the stroke of twelve.

And so she flees, with the prince in hot pursuit. She makes a clean getaway ... almost. One remnant of her magic hours is left behind at the scene of the crime (if deception be transgression in this context) -- a glass slipper which she lost while running from the palace just before midnight. The tenacious prince turns podologist, trying the slipper on every feminine foot he can find, for he believes that she whom the slipper fits will make the only queen fit to share his throne.

It is not too trying to turn this tale into an allegory of sorts. If a researcher has constructed a model corresponding to the interaction of organisms and their environment, it may prove difficult to judge whether a datum such as a phenotype or a particular behavior is relevant to fit or to fitness. Say that part of my model of prairie ecology holds that rabbits who run first and look later when they hear a threatening sound are more fit than those who look first to verify the direction of the threat and then flee. Every now and then a run-first rabbit will barrel right into the predator and be eaten. In those cases, the quickest of glances would have prevented such a mishap. But in general, there is higher mortality among look-firsts than run-firsts. I reason that this is because the prairie has few tall, solid objects, such as cliffs or large trees, which could reflect sound and thereby deceive the animal. (Obviously this ignores other sense data, the hunting methods of predators, and environmental features like deep-cut arroyos lined with cottonwoods, but even a contrived example will do for present purposes.) Thus I have a mental picture of fit and unfit rabbits.

Out doing field observations one day, I observe a veritable blitz of a bunny. The creature bolts at the merest sound of danger, never glancing to left or right but always moving immediately and at top speed away from the alarming noise. I tranquilize the rabbit, put a numbered tag on its ear, and confidently record in my journal that night: "Observed a very fit fellow between 7:30 and 11:45 this morning. Assigned him number 546 (purple tag)."

Now this leads to a distinction which echoes a point made when we reviewed the propensity interpretation. If I say that the blitz-bunny is fit, I do so because I associate speed with fitness. In fact I know nothing about that individual's reproductive history; for all I know, the rabbit in question might be sterile. Nor can I be certain that it does not possess a genotype associated with early mortality. Rather, I know merely that a single shoe fits, so to speak. It might fit any number of organisms, among them ones who are very fit, others who are quite unfit, and still others who are average with respect to whatever measure of fitness we are using (e.g., longevity or some way of characterizing reproductive success). In the fairy tale there was only one maid in all the land whom the slipper fit, but alas, evolutionary biologists don't have it so easy![36]

How do we find the "Cinderella level" at which fit and fitness correspond -- where a given aspect of our model is guaranteed not to collide with any other aspect of the model? The answer is that there is no sure way of finding this level. The best we can do is to constantly update an existing model to accommodate new data and new ways of looking at old data. Sometimes the modifications may be radical. Someday we might discover that look-first behavior had only coincidentally been associated with higher mortality. If we were to find as well that run-first behavior among prairie rabbits is half of a surprising pleiotropy (a case where more than one phenotype is associated with a single allele) whose complement is a new kind of rodent hemophilia, then our *prima facie* judgment of rabbit number 546's fitness would be less sanguine.

We have seen above that the process by which we update our models and thereby reckon fitness can be aptly described as recursive in the sense that a theory and its empirical support must be constantly revisited. (This goes not just for theories in which fitness figures, of course.) In fact constantly retracing our steps seems to be the only way of ensuring that we do not confuse fit and fitness. It may be objected that recursive means of analysis tend to be more complicated than iterative ones. Perhaps there is no means of generalizing as to which class of algorithms is simplest, but it seems clear that at least in some contexts a recursive method can be easier to conceive and execute than a non-recursive one, as we will now see.

### (4) Recursion as Leitmotif for creating systems of symbols and manipulating them

One need not have studied computer automation to conceive of the possibility that the same task can be perceived in numerous ways and that the way the task is approached will be affected by the mode of perception employed. A simple means of illustrating that fact would be to find an example of a procedure which can be completed serially or recursively. Again in the interests of simplicity, the recursion involved should depend on a single self-reference, for a total of two iterations. To put the matter another way, we are looking for something which has to be repeated just twice, and then it is done -- "it" meaning the task as well as the activities which, algorithmically, define the task.

Most readers of this chapter will have learned to tie their shoes long ago, but for some perhaps a memory of frustration will linger. A friend reports that as the youngest child in a large family, he was desperate to tie his own shoes. All of his siblings were privy to this bit of grown-up's magic, after all, and he wanted to be initiated as well. But the instructions seemed hugely complicated, even though the first part -- tying what sailors, boy scouts and other experts in the lore of knots call a "half hitch" -- was easy. What followed was a nightmare of complexity. The siblings nearest him in age explained the process with various metaphorical devices including one about a bunny running around a tree which was really a loop and then disappearing down his hole, which for the young seeker of knowledge was a mysterious region usually occupied by his clumsy, chocolate-smeared left thumb. (If you've never heard this bunny-round-the-tree story, just watch carefully as you tie your shoe. If you're like most right-handers, you'll make a "tree" loop in your left hand and then your right hand will pull the "bunny" around the tree and down the "hole.") After many days and not a few tears the lordly eldest sibling deigned to take notice of the theretofore fruitless training regimen and addressed his youngest brother with simplicity as well as majesty: "You already know how to tie a half hitch, and you already know what a loop is. Tie a half hitch with the plain shoe strings, as you've been doing. Then make each shoe string a loop and tie a second half hitch with the loops." (Again it might help to try this recursive algorithm yourself.) The knots my friend produced in this way were not works of perfect symmetry or tautness, but in abstract form they indeed represented *the* grail-knot which he had long sought to reproduce. And he rejoiced accordingly. End of story.

This homely example illustrates a phenomenon of concrete benefit and abstract fascination to software engineers: "Common programming languages such as Ada, Cobol, Fortran, PL/1, and C are all based on Turing machine theory; Lisp is based on recursive function theory....they are *very* different ways of looking at calculation" (Taylor 1988: 74). This last remark, about the different "ways of looking," applies not just to calculation but to evaluation of circumstances and to problem-solving in general. And it is not just that iterative, Turing-based and recursive approaches are different but of essentially equal utility. On the contrary, one approach may not "work," conceptually or practically, in a situation where the other method functions well, even elegantly. If we were to simplify and skip some steps, we could write one shoe-tying algorithm as follows:

ALGORITHM 1 ("Bunny Method")

(1) make half hitch; (2) make "tree"-loop with left hand; (3) make "bunny"-string with right hand; (4) "bunny" runs in front of the tree, then around it (going clockwise); (5) ensure bunny's "path"-string makes "hole," i.e., captures left thumb between "tree" and "path"; (6) just before having gone completely around the "tree," "bunny" dives down its "hole"; (7) tighten resulting loops by pulling them apart.

The meaning of "half-hitch" has not been specified, nor have exit conditions been spelled out. We assume that the user of the algorithm knows how to make a half-hitch and is likewise aware when the knot is finished.

For my friend in his younger years, the "bunny" algorithm simply did not work conceptually, however accurate it might be. What did work for him could be described simply as the "Half-hitch Method":

ALGORITHM 2: Half-hitch Method:

```
Procedure half-hitch(left, right).
        MakeLoop if (loop conditions).
        Until (exit conditions) half-hitch(left, right)
End
```

That's it. The boy simply makes half-hitch knots, with or without loops (depending on what stage the knot is in) until the knot is complete.

### (5) Generalizations in general

In the *Principia* (1913), Russell and Whitehead suggest that there are orders of generalization. If one says, for instance, "All generalizations are false," then the statement itself is not to be taken as belonging to the class of which the predicate, "false," is true. The statement is, in short, a higher order of generalization than the generalizations which it deems false. (This is part of Russell and Whitehead's general theory of types.) It is worth considering the possibility that some statements about fitness at large are similarly of a higher order than more specific claims involving fitness and synonymous terms.

If we say that as a general rule, the fittest organisms are those which survive, we may be taken as saying something like the following:

(1) Among all of the organisms which have been observed heretofore, those which have survived the longest were also the best adapted to their environments (i.e., the fittest).

Of course the definition could be fine-tuned to address reproduction and long-range survival of offspring, but for present purposes it suffices simply to consider individual survival. We can break this proposition down even further. Let the symbol /x/ denote the longevity of organism x measured in some convenient and appropriate units. For every set of organisms C: {a, b, c,...} in what we might call a "comparison class" (e.g., members of the same species in the same environment), "x is fittest" means:

(2) /x/ > /i/ for all i (other than x) in C.

It may be objected here that this definition of "fittest" could be replaced by some phrase (such as "longest-lived") which boils down to "survives the longest." In other words (the argument would go), a proposition linking survival and fitness is bound to be analytic.

But in light of Russell and Whitehead's theory of types, that conclusion is worth re-analyzing. It could be that only specific linkages of survival and fitness are truly analytic, whereas a generalization such as "Those organisms which are fittest are those which survive the longest" is in fact a higher-order generalization. In this case, that status means that fitness no longer includes the notion of longest survival. That would be a hard argument to make based only on Russell and Whitehead's theory of types for the simple reason that it is difficult to explain how the terms of a higher-order generalization can take on such different meanings that they escape properties (such as analyticity) which might inhere among the same terms in lower-order statements.

Our exposition of recursion in chapter three provides at least a partial response to this challenge. What we observed there was a difference between function and value based on Frege (1891). The distinction was not specifically bound up with recursion; rather, any statment such as A = f (B) comprises the conceptually separable elements of argument, function and value. This remains true in recursive cases, where A = f (A). There we can observe A either as value or as argument. Interestingly, A as argument can be expressed as a function of itself, too, as if it were a value, but at a lower level (or at a different moment of calculation, depending on the perspective). That "internal" self-referential expression will also have an argument that can be stated self-referentially, and so on, *ad infinitum*: A = f(f(f ... (A) ... )).

Now let us try to apply to the case of fitness what may amount to a version of Russel and Whitehead's insight that generalizations can have different orders. We observe two basic forms of statements relating fitness and survival:

(3) fitness = f (survival)

(4) survival = f (fitness)

Assuming that a garden-variety transitivity holds, we can then write not that fitness is a function of fitness, i.e.,

(5) fitness = f (fitness),

but rather:

(6) fitness = f ( f(fitness))

or fitness is a *function of a function* of fitness. This may seem to be splitting hairs if it is accepted that a "function of a function" simply amounts to another function. But there is at least an important rhetorical difference between statements (5) and (6) in so far as the latter underscores the extent to which fitness as argument is removed from fitness as value. They are at different levels of generalization, to use Russell and Whitehead's metaphor. The importance of this fact becomes evident when we consider that different Boolean truth values can apply at the different levels. For instance, Russell and Whitehead's generalization "All generalizations are false" can be read as true even though, as it asserts, any particular generalization is in some sense false (outside of certain contexts). In a similar way, the overall statement that fitness is a function of survival or that survival is a function of fitness ((3) and (4) above) can be taken as true without thereby implying that particular statements which appear as arguments of the function are likewise true. They may be false, even though they have the same form as the more general statement. In other words, although we can take the expression of recursive fitness below (7) to be true, some of the expressions of the same form which are embedded in the argument of the highest level generalization may have been falsified as hypotheses tested as particular statements (that is, as statements at the lowest level of generalization).

(7) $$\text{fitness} = f\,(f\,(f\,(f \ldots (\text{fitness}) \ldots))))$$

### 3. Appraising recursive fitness

Let us approach the matter in slightly different terms. Imagine we encounter the claim that a proposition linking the concepts of "survival" and "fitness" is deemed to be analytic (and therefore uninteresting) on the grounds that fitness is ultimately defined in terms of survival statistics. This allegation would claim, in other words, that fitness cannot be profitably employed in such statements because no *new* information is provided; if fitness is understood as a function of observed survival rates, then any proposition linking survival and fitness will have something like the form

(8) $$\text{present survival value} = f_1\,(\text{past survival values}) + \text{new data.}$$

which in turn can be expressed as

(9) $$\text{present survival value} = f_2\,(\text{past survival values})$$

But suppose we accept the general notion that one of the terms partially consists of the other, which is what the phrase "defined in terms of" seems to entail. Assume further that the *definiendum* (whichever word or phrase that happens to be) functions as a place holder: as the *definiens* changes across time, the *definiendum* is correspondingly updated. Under such assumptions we can imagine a progression of events at the beginning of which the *definiendum* and the *definiens* are precisely equivalent. In the terminology of set theory employed above, that would mean that the latter is an improper subset of the former. But that situation can be seen as an accident of time. As soon as the *definiens* is updated, the new *definiendum* will no longer consist merely of that current *definiens* but also of all previous definitions. In that sense, (9) is correct but can be stated more succinctly:

(10) $$\text{survival}_g = f(\text{survival}_c)$$

meaning that knowledge of survival in general is a function of current information on survival.

This way of looking at recursion in general and at survival in particular can be seen again and again in the literature, although I know of no instance where such examples are explicitly identified as recursive. Jared (1987), for example, summarizes recent findings which he reads as confirming Aristotle's and later Cuvier's contention that the number of teats in mammalian species is a function of the average litter size for that species. A survey of teat numbers and litter sizes had suggested that the rough ratio was 2:1, apparently a balance between efficiency (no more teats than the average number of individuals to be suckled) and safety (enough "extra" teats to accommodate larger litter sizes up to a limit which might be described as *fairly* common). The fact that human females have two teats corresponds to this insight: normal "litter" size is one, twins constitute "five percent of births in some populations...and triplets [are] vanishingly rare" (Jared 1987: 200). But Jared points out that there is a basic problem of interpretation: "This observed correlation begs the question as to cause and effect: does litter size determine teat number or vice versa?" The answer to the dilemma is provided by a simple observation: "within a few generations it is feasible to select animals for increased litter size but not for increased teat number, which is a rather constant species characteristic" (ibid.).

The analysis on which Jared based his reflections can be seen to have proceeded as follows. Organisms are fit in part if they can accommodate their offspring's survival needs with a reasonable expenditure of personal resources. "Reasonable" is here defined statistically in terms of the number of offspring produced. If the normal litter size of a given species is X individuals, if it is sometimes the case that the litter size is Y, and if it is rarely the case that a given species has a litter size of Z, then females of the species will tend to have Y teats. The meaning of "sometimes" here is determined by a repetitive analysis (one which *could* be schematized with a recursive algorithm): an individual is surveyed for teat number and litter size, then another individual is surveyed, and so on. At each point in the survey a tentative inference can be drawn based on the data provided by the survey as a whole. Any particular observation may correspond exactly to the substance of a given status of the inference (e.g., that the ratio of teats to litter size is 2:1) or not. That status, conceived as a propensity, may be taken as tentatively relevant to an overall definition of fitness (consistent with Mills and Beatty (1979)), but that status does not express the nature of fitness in an absolute sense.

### (1) Regresses, infinite and otherwise

Imagine two toy race cars, ones a child can wind up with a key at the back or side. The two cars set off toward a finish line with the exhortations of their respective owners urging them on. One -- the green Porsche, let's say -- finishes significantly ahead of the red Ferrari. The same race is run ten more times, and the Porsche wins all but one of them (when it crashes into the Ferrari at mid-course and gets the worst of the encounter). Let us say, following Mills and Beatty, that the Porsche is fitter than the Ferrari in this particular case because it has a propensity to finish first (demonstrated through empirical "behavior"). Now suppose we attempt to microreduce the alleged greater fitness to a specific quality of the winning toy (perhaps consistent with Weber 1996, depending on how one reads him). Maybe the tires are "stickier," getting better traction on the slick kitchen floor than those of the Ferrari, or maybe the Porsche's spring unwinds more quickly.

We can consider two broad results of such an investigation. First, we may look long and hard without finding that there is an obvious difference in the construction of the two cars. If that is the case, we may turn our attention elsewhere, wondering whether an outside factor may have influenced the "experiments." Could it be, for instance, that the Ferrari's owner did not wind his car tightly enough, or maybe he released it too slowly? If such is the case, then it would be more proper to say not that the Ferrari itself was less fit than the Porsche, but rather that the Ferrari-"driver" pair was less fit than the Porsche-"driver" pair. But once again we will want to know why: Is there a reason why the Porsche driver winds his car tighter or releases it more quickly than his buddy does with the Ferrari? If we answer the question, we risk adding another causal "layer" just as in the previous case. There we went from judging the fitness of a toy car to judging the fitness of the toy car plus another factor. That factor, once left out of consideration because it was held to be external, was then internalized: we inquired about the fitness of car-plus-driver. Now we are about to ask about the respective internal makeups of the drivers, and if we cannot find an obvious difference, we may add to the causal regress by examining the fitness of the car-driver-X triplet, and so on.

The second possibility is that we find an obvious difference at the current "level" of examination. For instance, we may find that the tires of the habitually losing toy Ferrari are hard and slick, providing significantly less traction than the tires of its rival. But it should be noted that even in such a case we *choose* whether or not to carry our questions about cause outside of the object's context. We might, for instance, choose to include the toy's designer or manufacturer in our questions about performance. Or we might bring the "driver" into the picture once more, asking why he chose such a toy in the first place (again making car-driver the level of examination) or why his parents chose that toy for a birthday present instead of a racer with better tires (in which case something like car-parents becomes the object whose fitness we must consider). It should be stressed that in this second case we *need* not expand the context of investigation, but arguably the option is available to us.

The conclusion is that causal regresses -- even infinite ones -- await those who choose to inquire at different "levels" of an experimental context. That may not be judged problematic, but it is a complication if one wants to use phrases such as "the fitness of that race car" or "the fitness of that car-driver pair" in a final, static sense, since we cannot be certain that fitness as something like Mills and Beatty's propensity exists *at that level* in a way which will satisfy our current investigative purpose. Causal regresses, in short, make it difficult to know whether "the fitness of x" is a meaningful phrase, that is, whether "fit" can accurately qualify an object at the level of x.

Now it may be objected that toy race cars provide a very bad analogy with organisms struggling to survive in nature. After all, the toys do not do what they do absent the involvement of external agents, and that scenario lends itself to the kind of "splitting" which results in fitness being applied to multi-level "objects." Let us see.

Consider the relationship between an organism and some fairly specific aspect of its environment. Reviewing the relationship of bees to the nectar of flowers, the entomologist Karl von Frisch noted that different kinds of flowers bloom at different times during the day. (He reports that Linnaeus went so far as to design a "flower-clock" by finding for every hour from 4 A.M. through 10 P.M. a species of flower which would open during just that hour. Alas, the flower-clock is not practically workable since the various species to be used are in bloom at different times of the year.) A less obvious kind of "periodicity" occurs as well: the amount and sugar concentration of a floral species' nectar change in the course of the day. Professor von Frisch concludes with blossoming metaphors: "Thus it occurs, that bees which fly to the flowers of a particular kind of plant will profit well at certain hours, while at other times of the day they will find the restaurant less hospitable or even completely closed" (1965: 256; my trans.).

Based on these observations we can imagine two hypothetical bees. One is familiar with the finer points of dining in the neighborhood, while the other is a bit like an American visiting Germany for the first time -- surprised to find many fast-food eateries already closed rather early in the evening by accustomed standards, and the grocery stores in the suburbs shut down for a couple of hours at midday on weekdays, most of the day on Saturday, and all day on Sunday. Such a bee never knows precisely when or where a meal is to be had. Not that it starves, but each sortie relies on trial and error rather than definite knowledge or instinct to achieve success. Now the savvy bee must not only know the optimal times to visit various flowers; it must also have some kind of internal clock so that it can "tell time." By the same token, the tourist may spend many a hungry hour because she lacks knowledge, clock, or both.

If we were to observe the respective behaviors of the two bees, it would seem natural to say that the savvy bee is fitter than the tourist bee, meaning that it is better adapted to its environment. That does not mean that we have somehow proved the existence of knowledge and clock in the first bee, and the absence of one or both elements in the second. Rather, our ethological observations show the savvy bee turning up at the right place at the right time so much more frequently than the tourist bee that we may infer what Mills and Beatty (1979) would have called "propensities" in the two bees, and we sense that the propensities may be reducible (or *micro*reducible, following Kim 1984 and Weber 1996) to some interaction among particular aspects of the environment and specific qualities of the bee in question. The savvy bee, who gains a full digestive tract with relatively little effort, would thus be fitter (have a greater propensity to long life and copious reproduction) than the tourist bee. Or so intuition or what we call common sense might lead us to believe.

But suppose we schematize the behavior of the bees in some fashion just to be sure what we mean by speaking of propensities. One way of doing this is to substitute bee *design* for behavior, so that the question becomes what a hypothetical, *optimal* bee would do (cf. Dennett 1995: 187 - 228 for a brief "philosophy" of engineering and its relationship to evolutionary biology). For instance, we might say that the behavioral *sine qua non* of the savvy bee is that it visits the right flower at the right time. (By that phrase -- *the right flower at the right time* -- we mean circumstances which an entomologist in the field might find extremely complicated, requiring detailed measurements of hive and flower locations, nectar amounts, sugar and pollen concentrations, and other variables. Those details need not concern us here.) But of course a single visit won't do. What we need to build is a bee which will visit the right flower at the right time of day, *day after day.* If the designed bee were driven by software, part of its code could look something like this recursive procedure:

```
PROCEDURE drink (FLOWER$_1$, FLOWER$_2$, ... FLOWER$_n$, DAY)
        <visit all the flowers in a certain order>
        DAY = DAY + 1
        drink (FLOWER$_1$, FLOWER$_2$, ... FLOWER$_n$, DAY)
END
```

In other words, the bee visits each flower at a certain time of day, then starts the same process all over again about 24 hours later. In the few lines of apian software above, the procedure is recursive -- it calls itself -- but such an algorithm *could* have been written in an iterative fashion. We could just as well have used a WHILE loop

(WHILE flowers remain unvisited DO visit them in a specific order) or an IF-THEN loop (IF flowers remain unvisited, THEN visit them in a specific order). But now suppose that the bee does nothing else but visit flowers. If this task is realized as a non-recursive function, then something will have to initiate that procedure each day. A programmer might call such a thing a "control module." It calls each procedure at the right time during the program's run. But in that case the bee's "software" would consist of that module as well as PROCEDURE drink. That means that the additional module could be more or less advantageous than the corresponding module of another bee. Moreover, we would need something to activate the control module, which in turn activates PROCEDURE drink at the proper time.

In a nutshell: a non-recursive explanation of an organism's overall fitness forces us to deal with an infinite regress as surely as a recursive one, only in a different way. We say that an organism is "wired" or "programmed" to behave in a certain way, but unless the wiring or program in question "calls" (activates) itself, then we will have to infer an activator function, which in turn will need activating. Put another way, if we consider the sum total of an organism's propensities to determine its fitness with respect to other organisms in the same environment, then to the extent that those propensities are recurring patterns of morphology (across generations) and behavior (intra- and intergenerational), the patterns must "activate" themselves; if they do not, then we will need to infer an activator function external to these propensities (contradicting the assertion that we had already taken into account all of the organism's propensities). The bit about morphology is important because above I have usually treated "propensity" as a tendency to *do* something -- to behave in a certain way -- but of course the notion of propensity applies to being as well as doing.

Let us come at the same point from a different angle. The notion of inclusive fitness functions essentially as a means of explaining apparently altruistic behavior. Another side of the same coin gets less attention but is relevant to the meaning of fitness nonetheless. This aspect might be considered a kind of nepotism, in which weaker animals benefit on a moment-to-moment basis (not just at moments of impending catastrophe) from their blood relationship to stronger organisms. Certainly we recognize this phenomenon in human circles; we wince or laugh at W.S. Gilbert's satirical view of the perks afforded higher officers in the Royal Navy, for instance, as the Right Honourable Sir Joseph Porter, K.C. B., First Lord of the Admiralty joins his relatives in song:

> Sir Joseph: I am the monarch of the sea,
> The ruler of the Queen's Navee,
> Whose praise Great Britain loudly chants.
> Cousin Hebe: And we are his sister, and his cousins,
> and his aunts!
> Relations: And we are his sister, and his cousins,
> and his aunts!
> Sir Joseph: When at anchor here I ride,
> My bosom swells with pride,
> And I snap my fingers at a foeman's taunts;
> Cousin Hebe: And so do his sister, and his cousins, and
> his aunts!
> Relations: And so do his sister, and his cousins, and
> his aunts!
> Sir Joseph: But when the breezes blow,
> I generally go below,
> And seek the seclusion that a cabin grants;
> Cousin Hebe: And so do his sister, and his cousins,
> and his aunts!
> All: And so do his sister, and his cousins,
> and his aunts!
> His sisters and his cousins,
> Whom he reckons up by dozens,
> And his aunts!

(1878: Act I)

But non-human animals -- chimpanzees, for instance -- afford their relatives similar protections (van Lawick-Goodall 1971: 130 - 3). In other words, the phenomenon is quite general. To repeat, it seems to amount to a milder version of the self-sacrifice which motivated the notion of inclusive fitness (q.v. Dawkins 1982: 185 - 186).

But how do we go about analyzing this kind of fitness? There will be *some* genetic similarity among the members of any given taxon, else the taxon would not have been conceptually established. We require a more definite thesis than that an organism will sacrifice itself for others *like* it. How much like it? That is the question which needs to be answered. If we say that an organism sacrifices itself or engages in any kind of nepotism for the sake of perpetuating its own genes -- a theory based on the work of Hamilton (1964a, b; cf. Dawkins 1982: 185 - 186) -- then we will also need to establish a cut-off point, a minimum standard of genetic similarity which will have to be met before an organism will ostensibly sacrifice any or all of its personal resources for the sake of another organism (cf. Dawkins' working definition of altruism in [2]1989: 4). We can choose to express that minimum as a function of the amount of its own genetic material which an organism would be able to pass on if it reproduced in a certain way under its own genetic circumstances (sexually in a diploid context, for instance). Notice that this common approach looks not just at reproduction itself but indeed at a significant aspect of behavior in general in a way which lends itself to recursive expression. If *fit* behavior is anything which propagates the maximum number of genes existing in the individual (or "vehicle," as Dawkins would have it) currently under consideration, then returning to the design metaphor we employed above with respect to the savvy bee, an engineer *might* code a portion of the organism's "decision software" in a recursive way.

What I think moves recursion from a *possible* tool for modeling how organisms make selfish or altruistic decisions to a *desirable* one is the goal of generality. Supposing we know enough about the principles of inheritance operative in a certain context, we may feel comfortable saying how a certain, simple organism will behave at a certain, simplified decision point where questions of altruism versus selfishness are relevant. Should an animal in a herd or a bird in a flock give a warning when it perceives a predator, even though doing so may draw attention to itself and therefore put itself in danger? If we grant the premise -- that warning endangers the signal-giver rather than having the opposite effect (e.g., by providing a cover of confused motion in which the individual can escape) -- then we can answer according to a selfish gene model: Yes, the warning signal should be given from the point of view of the signalling animal if it thereby probably can perpetuate its own genes more effectively than if it did not give the warning signal. But there are apparently many circumstances where we must *force* such an explanation by conceptually constructing the context in a certain way or by making other assumptions. Why would the Right Honourable Sir Joseph invite not just his sisters, but also his cousins and his aunts (who knows how many times removed?) aboard the ship on which he travels? Is it that when there are enough cousins and enough aunts whose survival chances are bettered by Sir Joseph's protection that he is as statistically likely or more so to propagate a certain percentage of genes like his own as if he had withheld his beneficence and lavished it instead on a more direct descendant? There can be no clear-cut answer until we gather much more data (perhaps more than we will ever be able to collect, depending on the complexity of the scenario); but to the extent that we can formulate an answer, the data will be gathered in what amounts to a recursive fashion as discussed in section 1.(1) of this chapter.

## 4. Conclusions

The most obvious question at this point is whether recursive fitness "advances the cause." In fact philosophers of science have two major goals when it comes to interpreting fitness. One aim is to understand how evolutionary biologists actually use the concept of fitness in their work. The second purpose is to argue that propositions involving fitness can succeed in a logical sense, that is, that they can avoid circularity (interpreted, e.g., as analyticity). It will be recalled that Mills and Beatty (1979) claimed success for their propensity interpretation on both counts.

We have already summarized and reviewed Mills and Beatty's well-known propensity interpretation of fitness (1979). Further, it has been suggested that a recursive understanding of fitness provides the best means of understanding how fitness logically functions in propositions which are of real interest to evolutionary biologists. Under these circumstances the question naturally arises: Isn't the propensity interpretation an essentially recursive understanding of fitness? After all, we reckon propensities by the repeated observation of happenings which are the same (or nearly so) in circumstances which are the same (or nearly so). Want to know what propensity a coin has to turn up heads? Then start flipping coins, first making sure that each trial is nearly enough similar to all the others to be of comparative value. It is not hard to meet that criterion in the case of coins, since not many have obvious abnormalities (magnets taped onto one side, for instance), nor do most flipping environments differ from one another in ways relevant to the average trial (magnetic plates on the table with polarity opposite that of the magnet taped to the dubious coin). But in other cases, extreme care is needed to ensure that trials really amount to repetitions. Assuming we can construct sufficiently clever experiments or perform exact enough observations in the wild, we evaluate propensity by doing essentially the same thing over and over.

It has been said that when one's only tool is a hammer, then every problem looks like a nail. The notion of recursive fitness, as it has been developed thus far, does not deny the utility of the various interpretations of fitness discussed. To repeat what has been said earlier in different ways, the aim of recursive fitness is to supplement other interpretations, not necessarily to replace them. It would be a mere side-effect if recursive fitness as presented here can also subsume the other interpretations by taking them as particular, specialized instances of a generality expressed recursively. It cannot be concluded from that broadly encompassing nature that recursive fitness denies the worth of the other interpretations any more than one would say that a broad theory in any realm of human inquiry or activity renders useless the theories which it generalizes. Some have argued, for instance, that Hinduism *per se* does not exist; instead, it is a name given for conceptual convenience to a broad range of faiths which mutually recognize one another as legitimate means of striving toward what are, very abstractly understood, the same goals.

> So elusive is this ancient and cumulative religious tradition that some scholars have despaired of definition and suggested that Hindus are identified simply as the religious remainder after one subtracts all Muslims, Jainas, Buddhists, Christians, Jews, Parsis, and tribals from the religious landscape of South Asia. (Knipe 1991: 1-2)[37].

Other scholars argue whether Jainism is a part of Hinduism (Matthews 1991: 187). That leaves a huge variety of beliefs and practices which in some broad sense must be specific realizations of the umbrella concept.

The analogy of recursive fitness with Hinduism can be carried too far, of course, but it serves to indicate a basic strategy of the argument to this point. The goal is not so much to destroy the conceptual options presented in, say, the propensity interpretation of fitness as to engulf it -- a “logophagy” intended not to weaken earlier interpretations of fitness but rather to strengthen and justify the concept of recursive fitness by making it broader than, say, straight actuality or propensity interpretations. But of course there must be *something* to be added to these earlier discourses on fitness, else the arguments presented in this dissertation would be wholly unnecessary. That something is range. It has already been suggested that the propensity interpretation devised by Mills and Beatty (1979), for instance, bites off more than it can chew.

It may well be objected that fitness cannot be treated as recursive, since that qualifier (the objection would continue) applies to algorithms rather than objects or things. It would then be further claimed that fitness has the status of a thing. (This thing may be dubbed a quality, a propensity, a facet of an organism’s overall being, or call it what you will. The point is that received interpretations of fitness emphasize quiddity rather than method.) But we have seen that recursion can be conceived both as object and as method. In fact, the notion of recursion does not just straddle the two realms of object and method; it blurs them. As Wirth's presentation shows -- and in this respect his exegesis is representative of the way recursion is treated by computer science in practice as well as theory -- to speak of recursive method is to have in mind recursive objects as well, and vice versa. Perhaps that is not so surprising when we reflect that many "objects" of science are inferred from specific procedures. The difference between such objects and recursion is that recursive algorithms constitute not so much a specific procedure as a category or pattern of methods. Moreover, it seems unlikely that recursive objects are necessarily "discovered" by recursive algorithms. But the two -- object and method -- are bound up with one another in the case of recursion as well as in that of some scientific methods and the objects inferred from them.

Of course a close connection might be found to exist between *any* comprehensive description and its object. The special character of recursion lies in its non-linearity, its ability to turn back on itself. This is opposed to a quality which has been variously described. One way of looking at most verbal descriptions, for instance, holds them to be "discursive" or what we might call "projective." Langer exploits these terms (having borrowed the notion of projection from Wittgenstein's *Tractatus*) in an exposition of what she takes to be the *non*-discursive, *non*-projective nature of symbolism in much of art and ritual ($^3$1957: 79 - 84).

This dissertation began with analogies such as the following:

ancient engineers dealing with arches: calculus :: modern biologists dealing with fitness: recursion.

The point is that fitness as a concept can be accommodated by a recursive model, but is that model necessary? It may seem that the thrust of the argument to this point has been to answer in the affirmative. But in fact the answer is probably negative: No, fitness is not constrained in such a way that only recursion can describe it, so far as we know. Something better may come along. To the extent that recursion describes a hypothetico-deductive process which we feel is necessary for science in general, however, there does seem to be a certain necessity in the concept of recursive fitness. We will test this further in the next part.

# Part Three: The Locus of Fitness

# Chapter Nine: Normality

This chapter and those following further explore the notion of recursive fitness at two levels corresponding to the questions "What specific things can be fit?" and "What is fitness in the abstract?" We will also review what I take to be the most important interpretations of fitness while at the same time repeating a basic critique of each position. Fitness has figured in numerous writings in evolutionary biology and related fields, but the meaning of the term has seldom been carefully "unpacked." A practicing biologist, convinced that he and his peers already understand the term, might well feel no compulsion to expound its meaning at great length. But this does not mean that the concept is unimportant in evolutionary biology. "Fitness" is not an entry in the index to the facsimile edition of Darwin's *Origin*, though no one would dispute that the concept is central to Darwin's argument. Even some of those who recognize the difficulties in defining fitness and how it figures in evolutionary biology as a whole prefer to pursue their researches without pausing to tackle the "fitness problem." In 1976 Dawkins "avoided using the word [fitness] in this book" because a certain kind of mistake "is all too easy to make when we use the technical term 'fitness'" ($^2$1989: 137); later he took pains to show that use of the term does more harm than good (1982: 179 - 194). As we have seen to this point, there have been more optimistic philosophical attempts to elucidate the role of fitness in evolutionary biology with the aim not of discrediting all talk of fitness, but rather in order to rehabilitate our understanding of the concept. The best known of these attempts, taken together, form a convenient foundation for further study of how the concept functions. This "canon" of fitness studies includes not just claims to be scrutinized but also a handy vocabulary which we have appropriated for our present purposes.

Theories in evolutionary biology, including those which employ the notion of fitness, are about entities of various sorts -- sometimes individual organisms, at other times phenotypical facets of individuals such as organs or skeletal structures, and still again about groupings of organisms such as species or genera. At each one of these various levels certain qualities are of interest to researchers while others are scarcely acknowledged. A necessary first step in understanding scientific thought about these entities and qualities in general, and about their relative levels of fitness in particular, is to grasp how and why they are abstracted from the soup of nature to become delineated objects of study. What is it that makes the blood-pumping aspect of a heart more interesting to scientists than some other of the organ's properties -- its smell, say? Why are organisms which clearly differ in their basic structure nonetheless grouped together for some purposes? Perhaps because they are part of an alleged genetically-describable lineage? In tackling such questions as these I do not propose to revisit well-known debates like those between cladists and morphologists among taxonomists, but rather to focus on a tool common to all the various camps engaged in all the sundry discussions on how to cut up and classify our observations of the living world. That tool is the concept of *normality*, and by understanding how this notion is or could be used in certain contexts of evolutionary biology, perhaps we can approach the tough subject of fitness with more confidence and a greater hope of success.

To that end a particular understanding of normality as the foundational notion underlying the establishment of taxa must be developed. A new perspective on debates over fitness (e.g., whether propositions linking fitness and survival are tautologous) can be had by understanding the problems at a more basic level -- the level at which organisms are classified. Ultimately this perspective may lead to a view of fitness which differs from what one might call the common-sense understanding (e.g., in Caplan 1977), the propensity interpretation (Mills and Beatty 1979) and from Sober's 1984a and 1993 explanations of the concept. None of these concepts of fitness is explicitly contradicted in this chapter, and in fact only the propensity concept is discussed at any length. Instead, the focus is on normality, a concept which is seen to depend on three factors: the organizational principle upon which the taxa under consideration are established, the way in which individuals are counted, and the reckoning of probability which underlies the understanding of how organisms of a given lineage can change across generations. This argumentation is similar in form to Sober's (1984a, 1993), but important differences will be seen.

Corresponding to this triad of factors -- ways of organizing, counting and calculating, one might say -- three discussions are developed as a groundwork.. First, normality is seen to be the key concept allowing the delineation of natural taxa. Wachbroit's 1994 discussion of three kinds of normality is the "straw man" for this way of proceeding. What we can call "normative" normality (Wachbroit's "evaluative" normality, akin to moral prescription) is discarded as uninteresting for present purposes; Wachbroit's "biological" normality is shown to be a derivative of what he calls "statistical" normality. Diverging even further from Wachbroit, three interesting cases of statistical normality are then defined and examined: "delimiting," "circular," and "F-" normality (the latter named after Frege's 1879 account of how equality may be established without counting). These different ways of interpreting normality bear on the ways we can understand fitness.

The second discussion -- on means by which we can count organisms and their traits -- emerges naturally from the reflections on normality. By seeking an account of how something can be termed normal without explicit reliance on counting in the sense necessary to undertake statistical calculations, one can imagine how taxa might be conceived based on F-normality. If this is so, then apparently we face both a problem and an opportunity. The problem can be expressed as a question: If changes across generations of a taxon are describable only probabilistically, how can we describe the evolution of organisms which are not rigorously counted? Developing a response to this question allows us to reevaluate what we mean when we use the terminology of probability to describe the phenomena of evolution.

Naturally this reevaluation requires a short discussion of probability piggybacking on what has already been said. The result of this investigation matches Sober's view to this extent, that definitions of fitness are closely tied to commitment to one school or another of probability. To the extent that the propensity theory conceives fitness propensities as finished, static properties of individuals, it dovetails most neatly with a frequency account of probability, whereas Sober's 1984 insights depend more on a range theory. Further, what is here called the "common sense" account requires clarification at a foundational level -- meaning the level of probability theory -- before it can be considered an adequate evaluation. Key to this clarification is consideration of a "hard case"-- a feature which cannot be accounted for genetically (i.e., no specific genes have been correlated with the factor) and which also cannot be entirely explained by appeal to morphology. Human capacity to use language seems to be such a factor. That is, human languages evolve *somewhat* as whole organisms do, though the extent remains to be seen (cf. Lewontin 1974 in Sober 1984c: 5; Futuyma 1983: 153, 160). We will wait until the final chapter to consider in detail an aspect of language as an evolutionary phenomenon.

None of the understandings of fitness mentioned above adequately accommodates the full range of possible organizational principles, counting methods, and probability theories which provide the foundation for any general account of evolution. On the other hand, each interpretation of fitness performs admirably in its own specialized region of application. The chapter ends by repeating that a recursive model of fitness can accommodate the key aspects of our conception of normality.

## 1. Kinds of normality

When we say that an organism belongs to a given taxon, we are claiming that certain of the organism's qualities are within specific parameters. When we say an organism is *normal*, we are presumably saying the same thing. In other words, the concepts of normality and membership within a taxonomic group bear some very intimate relationship to one another. This very general acknowledgment of the primacy of normality in taxonomy seems self-evident. Only when the nature of the parameters themselves is discussed would one expect to see hackles raised as various warring camps -- essentialists and nominalists, for instance, or cladists and morphologists -- pursue their agendas. Investigating normality at a very basic level, one which precedes the agendas of these camps, is a prerequisite to further speculation on the nature of taxa and how evolutionary theories may posit their existence as taxa and speculate about their development. The discussion which follows will pay particular attention to normality as it bears on questions of taxonomy and evolution. Failure to understand what we mean by the term "normal" as applied to organisms in these contexts may leave us at a later point facing the obscuring fog which Elliott Sober describes in the opening paragraphs of his *Philosophy of Biology* (1993).[38]

The concepts of probability and randomness clearly play a large role in the Darwinian account of evolution. Much of evolutionary theory necessarily looks backward from the current condition of a given group of organisms and employs notions of probability to demonstrate how previous groups could have become what we see before us. By either a cladistic or morphological grouping of organisms, taxa (e.g., species) are themselves established based in part on judgments of probability. A difference of one centimeter in the measure of some feature of an animal species may be considered so *normal* (i.e., *likely* to be observed) that its belonging in that particular taxonomic grouping is unquestioned, whereas a difference of two centimeters may mean disqualification. Classifying things on the basis of likelihood or normality is of course not peculiar to evolutionary theory but pervades scientific disciplines and practices in general. Moreover, normality as an observational tool is employed to judge behavior (process) as well as structure. Evaluating behavioral phenomena as normal or abnormal may in turn lead the researcher to reductively attribute unobservable structure to the organism in order to explain the behavior. Psychiatrist Oliver Sacks, for instance, describes

> ...cases of the obviously pathological -- situations in which there is some blatant neurological excess or deficit. Sooner or later it is obvious...that there is 'something (physically) the matter' [with certain patients]. Their inner worlds, their dispositions, may indeed be altered, transformed; but, as becomes clear, this is due to some gross (and almost quantitative) change in neural function. (1985, 129-130)

Presumably no one would care to argue that mentally abnormal patients are non-human, and indeed frequently abnormal behavior can be accommodated within the concept of a given species (e.g., in the example of *Chrysopa downesi* and *carnea*, which *can* interbreed, as proved in the laboratory, but which do not do so in the wild and therefore conform to Mayr's concept of a species as a population which does not naturally breed with other populations). But in other cases behavior is crucial for determining how to group organisms. Generations of scientists have defined species in part on the basis of ability to mate successfully and, until genetic explanations became common, simply inferred some common structure as the cause of such mating behavior.

So we speak of normality with respect to structure and behavior, and these aspects of organisms determine, at least to some degree, the taxa we discern. But normality in both realms is apparently closely tied to what is most frequently observed, that is, to what is probable. In other words, what is *normal* in our mundane experience as well as in our careful observation of the natural world -- what occurs "always and for the most part" in Aristotle's phrase -- is arguably the same as our concept of what is *probable* and the opposite of that which is either seldom seen (the *improbable*) or which defies prediction (the *random*). These concepts of probability are in turn crucial to our understanding of Darwinian evolution as a process of genotypic change in an environmental context where the inheritance of randomly occurring traits can be assigned a quantitative probability.

If all of this is granted, we have an especially strong motivation to know precisely what "normal" means in contexts related to the theory of evolution. Does this notion have to do only with frequency, with counting up the occurrences of phenomena as we experience them? This is what Aristotle's phrase suggests: *always* or *frequently* observed implies normality; *seldom* or *never* seen implies abnormality. Or is it possible to conceive of something as being "normal" without having to take a numerical survey? Robert Wachbroit (1994) argues that there are two alternative conceptions of normality besides what he calls the "statistical" -- namely the evaluative and the biological senses of the term.

### (1) Normative normality

It seems clear that Wachbroit is correct to this extent: a legitimate understanding of "normality" could certainly rely on something other than frequency as its basis. Common usage of the term shows that its sense can indeed be given by moral (what we might call "normative" or what Wachbroit terms "evaluative") considerations which ignore actual frequency of occurrence in order to further an ulterior, morally prescriptive agenda. Both *normal* and *normative* derive from Latin *norma*, meaning a carpenter's square, and thus it may seem redundant to speak of a "normative sense of normality." But perhaps the odd phrase "normative normality" will underscore by contrast the meaning of our primary focus in this section, namely "normal" as "frequently occurring." One would expect that in most aspects of life, actions which are considered morally bad are in fact abnormal in the statistical sense -- few of us commit murder or robbery in the accepted legal senses of those terms. Thus if one says that committing murder is abnormal with the intention of implying that such an act would be morally wrong, it might still be claimed that this usage is consistent with the statistical sense of the term as well. But of course we might find exceptions in which what is called normal or abnormal in a normative sense violates the corresponding statistical sense. Our culture's most sensational (and therefore explanatorily useful) employment of normality in a morally prescriptive sense probably has to do with sexual mores. An example from a text used to teach medical students of the 1960s how to evaluate physical condition of patients asserts: "Hypertrophy of the clitoris suggests endocrine dysfunction or *abnormal* sex behavior (masturbation)" (Prior and Silberstein, 1963: 299; emphasis added). Surveys of homoerotic practices as well as physicians' common sense together with their knowledge of scores of patients presumably would have made it clear to the authors that the behavior under consideration could not be called statistically abnormal; instead, the sense in which they use "normality" here is normative. (Perhaps the doctors who authored the text were unaware of any actuarial data and could not draw any contrary conclusions from their professional experience and observations, but it seems more probable that their attribution of *abnormality* in this context is instead their homage to the moral norms of their generation and of their profession.) These examples (which do not appear in Wachbroit's work) suggest that there is indeed a sense of normality which is not statistical.

For our purposes in this section, however, the morally laden sense of normality is unimportant because its *primary* goal is rhetorical rather than scientifically probative.[39] What we seek is an exhaustive description of scientifically relevant normality so that we can confidently undertake a rigorous analysis of Darwinian evolution. The salient question therefore becomes: Is there any non-normative sense of normality which is relevant to biological phenomena? If so, then our speculations about evolution must use the concepts of probability and normality in such a way as not to confuse the different senses of these terms.

### (2) Kinds of statistical normality

Wachbroit develops a concept of normality which he believes differs from what he calls "statistical" as well as "evaluative" (morally normative) normality. He calls this third and allegedly different understanding of normality "biological." By Wachbroit's account biological normality differs from the concept of statistical normality which is employed in the physical sciences, although at the end of his presentation he cautions that "[t]his discussion of the distinctiveness of the biological concept of normality does not *prove* that the concept cannot be analyzed or reduced into the language of the physical sciences..." (590). It is hard to know what to make of this caveat, since Wachbroit clearly aims to present a non-quantitative concept of normality. Moreover, he believes that this non-quantitative brand of normality is frequently used in biological sciences.

> Consider one of the favorite examples of the philosopher of biology, "the function of the heart is to circulate blood". That statement is clearly not about any *particular* heart (that is, no particular heart is named or intended). Nor is it a statement about *all* hearts, since some hearts (sometimes) fail to circulate blood. The statement is not even about *most* hearts or about the *average* heart. Suppose a calamity occurred in which most people's hearts failed to circulate blood so that they needed an implanted medical device for this purpose. This would hardly undermine the statement about the heart's function. Nor, finally, is the statement about the *ideal* heart. What the statement *is* about is the *normal* heart, with the understanding that a particular heart or all hearts or most hearts or the average heart or the ideal heart may not be normal. (580)

The implication here is that we can at least sometimes speak of what is normal without counting occurrences of the phenomena under observation. Wachbroit does acknowledge that "statistics, for example, may provide important evidence for determining biological normality and biological functions" (581), but whatever he may mean by the phrase "provide important evidence," it is clear that he wishes to divorce biological from statistical normality. Instead of counting for the purpose of undertaking rigorous statistical analysis and then establishing what is normal, by Wachbroit's account we base our notion of biological normality on something like an appreciation for proper functioning within a system.[40]

An obvious reaction to this claim is to question the *origin* of any such non-quantitative understanding of normality. Must we not at some point in the past have done our counting in order to understand both the context (circulatory system) and the function (pumping blood) of the object under consideration? It may be true that we can at present conceive of a normal heart without conducting a large-scale survey of human hearts in general, but presumably this is only true because we have in some way inherited the results of such an investigation. The observations and the counting of their frequency may have been carried out formally in the recent past, but certainly we are also the beneficiaries of what past generations of anatomists, surgeons and others ranging from Leonardo to Harvey to amateur pathologists have observed and passed on to succeeding generations. If this is all we mean by "biological" normality -- a piggybacking on past statistical observations -- then any claim of a clear distinction between statistical and biological normality would seem to be exaggerated. (Wachbroit appears purposely to avoid making such a claim, but he also does not explicitly consider this response to the distinction he wishes to draw.) Absent Wachbroit's specific response to this challenge we can speculate no further on how different he perceives biological and statistical understandings of normality to be. But perhaps we can take his claim that there is a real sense of normality which is neither normative nor purely statistical as a challenge to examine the concept of statistical normality more closely. Our aim will not be to posit a non-statistical understanding but instead merely to ensure that we do not unknowingly employ two or more essentially different notions of normality -- all of them essentially statistical -- in our analysis of fitness and of evolutionary theory in general. The concept of normality is after all so basic, so foundational, that any confusion at this level could threaten whatever theoretical edifices are built upon it.

### (3) Normality based on counting

In essence the concept of statistical normality is based on counting and comparing. Count up all the thingamajigs sporting Xs, then count up all the thingamajigs featuring Ys. If there are lots more Xs than Ys, then presumably it's normal for thingamajigs to have Xs. If the numbers are close, we may have to increase the sample space. If the numbers remain close even after we have significantly expanded the range of our observations, we may decide that it is normal for thingamajigs to have either Xs or Ys. The following statement by Stephen Jay Gould, for instance, can be read as claiming that *normally* life on earth is bacterial rather than vertebrate:

> The most salient feature of life has been the stability of its bacterial mode from the beginning of the fossil record until today and, with little doubt, into all future time so long as the earth endures. This is truly the 'age of bacteria' -- as it was in the beginning, is now and ever shall be. (1994: 87)

This judgment is based on simply counting exemplars of different life forms, estimating the size of the population each represents, and then comparing these numbers to the number of living things overall. The ascendancy of bacteria as the normal (most frequently occurring) kind of life allows startling comparisons: "The number of *Escherichia coli* cells in the gut of each human being exceeds the number of humans that has ever lived on this planet" (ibid.).

Calling bacteria the *normal* sort of terrestrial life seems perfectly consistent with a notion of statistical normality (rather than "biological normality" à la Wachbroit). At the same time, saying something like "Bacteria are the normal lifeforms on earth" seems oddly awkward, something one would not say. Instead we would be more likely to assert that "Bacteria are the *most common* or *most frequently occurring* life forms." Perhaps "normal" is usually reserved for comparisons to a mental archetype, a possibility we will consider below. For the time being suffice it to say that there does seem to be a sense in which what is "normal" is simply what occurs with the greatest statistical frequency. Further, when we view usage of the term from a linguistic and not just a purely numerical standpoint, it appears that judgments of normality rely on groupings of terms: conceptually, one unites two concepts in a kind of mapping. But an account of statistical normality as something defined by the frequency of such mappings is not unproblematic.

## 2. Two problems

### (1) First problem: Delimiting statistical normality.

The first problem has to do with how we decide what to count in order to build a statistical foundation of normality. Three versions of this problem present themselves, and each can be stated in the form of a claim: (1) Metaphysically (ontologically) speaking, there are no shared features of organisms which allow them to be grouped; (2) We have no access to whatever regularities might exist in nature, and therefore we lack the epistemological foundation to construct a taxonomy based on perceived normalities; (3) We have access to regularities upon which we can base our determinations of natural kinds, but we have no firm standards for choosing one locus of regularities over another when we "slice up" nature.

The first two claims, each consistent with a radical nominalism, seem uninteresting in this context. The third claim is not only interesting but indeed stands at the crux of ongoing debates over taxonomy (based, e.g., on cladistic versus morphological schemes). Certainly this claim deserves close scrutiny.

Count up all the hearts within certain parameters and compare that number to the number of hearts outside those parameters. Based on this comparison one will presumably learn what constitutes a normal heart. But where do the parameters come from? The answer is (of course) that the counting is done in conjunction with analysis of the object counted. Naturally one would expect the investigation of an organ such as the human heart to be conducted over many generations; thus what physiologists can tell us about the human heart is certainly much more than anyone knew just twenty years ago, but their knowledge is nonetheless indebted to ancient as well as modern sources. Presumably at every point in this long investigation the study was directed by theories as to what role the heart plays in the human being (its function) as well as how its activities complement and are complemented by the workings of other organs (also essentially function-oriented observations). This influence would naturally lead researchers to focus on some aspects of the heart (e.g., size, number of ventricles) and ignore others (odor). Because the evaluation of normality depends on parameters which are determined at least in part by speculation about function, normality can itself be seen as a functional concept: we may decide to ignore entirely something that is common to almost all hearts when we establish the parameters of normality. But there is a risk here. What if we have failed to notice some unifying or distinguishing feature among the objects or organisms we are examining? What if we disagree (as cladists and taxonomists of other ilks do) over the significance of the structures we observe in existing organisms? Sober notes that

> In the end, the fatal flaw in evolutionary taxonomy is that it has never been able to formulate a nonarbitrary criterion for when homology matters more than propinquity of descent. Crocodiles and lizards share a number of homologies that birds do not possess, but it also is true that crocs and birds share homologies that distinguish them from lizards. Why are the former homologies so much more important than the latter that we should decline to classify by propinquity of descent? (1993: 168)

Here Sober is addressing a more specific matter than the issue of how we go about determining which criteria are relevant to our judgments of normality. Nonetheless, his point is applicable to the more general question: What is our standard of abstracting qualities for analysis and comparison?

Also for a somewhat different purpose than answering this question, Gould and Lewontin (1979) express perhaps the strongest, most far-reaching doubts about the process of selecting traits of organisms and then subjecting these isolated features to rigorous analysis. They argue for a holistic conception of organisms primarily in order to avoid misleading inferences that individual characteristics result from environmentally motivated adaptation, but extrapolation from their speculations leads to general doubts about the grounds of choosing isolated qualities to determine what is normal. In Gould and Lewontin's account the issue is not so much that the researcher may be paying attention to the "wrong" feature of an organism, nor is the criticism that selection of qualities as candidates for analysis is arbitrary and therefore an inevitably shaky foundation for far-reaching theories; rather, the concern is that some aspects of organisms in themselves represent nothing about evolutionary history. By this way of reckoning some qualities are simply forced responses to other, primary aspects of the organism's overall *Bauplan*, just as the spandrels seen in the cathedral of San Marco in Venice are not intentional elements of design but instead are "necessary architectural byproducts of mounting a dome on rounded arches" (in Sober 1984c: 253). (Gould and Lewontin make it clear that they consider physical structure to be the realm of spandrels -- structures which are necessary given other aspects of an organism's *Bauplan* rather than being the results of adaptation -- but perhaps phenotypic qualities are not the only possibilities. Although mappings of geno- to phenotypes may not be one-to-one, a group of genotypes may be seen as the necessary accompaniments of a given phenotype-qua-spandrel. Looked at in this way -- as necessary accompaniments of more broadly conceived macrostructures -- perhaps genotypes can also be conceived as forced in a backward sort of way.)

Dennett (1995) criticizes the anti-adaptationist message here by suggesting that what Gould and Lewontin mistakenly call a "spandrel" is by no means a necessary byproduct of the dome-on-rounded-arches superstructure, since pendentives (said to be the proper term for the structural element under consideration) could just as well have been omitted in favor of a much different architectural form (e.g., a squinch) than what the cathedral's designers finally settled on. The obvious question here is whether Dennett is stubbornly refusing to acknowledge a real design necessity, which although falsely attributed at the level of pendentive and squinch, is nevertheless real at the once-removed level of spandrels in general. Surely one among a limited class of forms must fill the space created by the overall structures. Dennett seems to acknowledge the possibility, but he quickly rejects it.

> But is there nevertheless some other way in which spandrels in the narrow sense -- pendentives -- truly are nonoptional features of San Marco? That is what Gould and Lewontin seem to be asserting, but if so, they are wrong. Not only were the pendentives just one among many *imaginable* options; they were just one among the readily *available* options. (1995: 273)

Thus Dennett asserts that Gould and Lewontin are wrong in their divergence from a stricter adaptationism. His alternative to "*Bauplan* thinking" is a thoroughgoing belief that structural features of organisms reflect natural selection of advantageous traits. It is the selective advantage, he claims, which explains the choice among available options.

Any synthesis of these two positions (assuming one is to be found) presumably would have to rely on the notion of optimization. If a given environment seems to make one design choice among alternatives the *best*, then analysts may fairly infer that the choice was made adaptively. In so far as *some* choice had to be made due to other aspects of the overall *Bauplan*, however, the alternative actually chosen is in some sense also *necessary*. But an important question remains unasked: What is the *Bauplan*? What constitutes the part of the organism which determines where spandrels must be filled in? The only possible answer would seem to depend on design *chronology*. Elements of structure which come first make up the *Bauplan*, while those that come later are dependent on this original structure. But that answer is not wholly satisfying in the context of Gould and Lewontin's metaphor. The cathedral's designers may well have known they were going to produce the pendentives we actually see before they laid the first stone. It is *possible* that these architects would have sacrificed the macro-structure if it had not allowed the pendentives. Similarly, it is possible that what we think must be the primary structure of an organism is in fact secondary to its "spandrel"--an apparently less central portion of its morphology. In the case of the cathedral, this may seem to be a remote alternative. Based on Gould and Lewontin's own support of holistic appraisal of organisms, however, is it so hard to imagine that the designers would have disapproved the entire design if the pendentives had been squinches? An historian of architecture may have good reasons for believing so, but there seems to be no necessary, *a priori* reason. Similarly in the case of organisms, the function of a given structure can be perceived in different ways and the structure itself can be assigned varying degrees of primacy apart from its place in the historical development of the overall organism. Futuyma illustrates the concept of "preadaptation"-- "a structure that takes on a new function"-- by appeal to the beaks of parrots. Many of these birds are in fact not carnivorous, but because the New Zealand parrot *Nestor notabilis* "rips through the skin of sheep with its sharp beak to feed on fat..., many parrots could be said to be preadapted to carnivory" (1986: 424).

What this all goes to show is that analyzing organisms in terms of primary *Bauplan*, necessary spandrel, and adaptively selected pendentive is not a sure solution to the problem of how we select features of organisms on which to base our notions of normality. If we observe "pendentive" organisms and "squinch" organisms, should we count them as different, or should we ignore the micro-design in favor of observing the overall *Bauplan* (dome on four columns, or vertebrate, say)? In fact even the most apparently "unforced" aspect of a proposed *Bauplan* may be explained as the result of broader constraints, assuming we are interested in entertaining reflections on the designer's motives. Thus anything other than a steeple of a certain kind or a dome of a certain shape may be unthinkable to a pious designer whose creative abilities are the product of his culture and beliefs. Similarly, any *interesting* micro-feature of a *Bauplan* may be seen as a choice. (Here "interesting" means "not wholly determined by physical laws." The fact that a roof is supported by *something* -- walls, columns, arches, etc. -- is uninteresting because of the existence of gravity.)

One way of characterizing the adaptationist/bauplanist debate is as a question about significance: What can any facet of an organism *mean*? What can we conclude from it? To highlight this facet of the debate, consider what something like a phenotype signifies under one theoretical framework as opposed to another -- say a theory of the emergence of diversity which permits elements of design as opposed to a theory which is wholly mechanistic. In their discussion of what Dawkins' concept of the extended phenotype entails, Hampe and Morgan (1988) have chosen not to stress what I take to be the most dramatic consequence of Dawkins' ideas as they may be applied to human beings, particularly in an era when we are on the verge of engineering ourselves genetically or even augmenting ourselves with non-organic technologies (Minsky 1994): a blurring of the distinction between artificial design and organic growth. Or at least Hampe and Morgan do not use these terms in their critique, though they seem very close to doing so at times: "Dawkins can be interpreted as wanting to remove the distinction between actions that are explained intentionally and organic processes that are explained physiologically and -- in the end -- genetically" (1988: 130). One wonders what Paley would have concluded had he found not a watch in the sand, but rather a human being with a Jarvis heart, engineered genes, and microchips embedded in the brain and neural pathways. Would the find be evidence of evolution, design, or both? Even if the capacity for design as a kind of phenotype comes ultimately from a mechanistic process of evolution driven by genes, the fact of design and the intentional state underlying the capacity seem undeniable.

We can make the same point about significance in yet a different way. Dawkins' notion of the meme (Dawkins 1976/$^2$1989: 192) as "atom" has counterparts in other disciplines. This is what we would expect. Scientists attempt to abstract the building blocks of natural phenomena according to different standards. Frequently it is best if the building blocks themselves are irreducible, or as nearly so as possible. (If the units in which a scientific theory is framed are themselves reducible, then the researcher may well feel that the theory is not expressed as *generally* as it might be. This is because the "sub-units" must be responsible for any phenomenon expressed in terms of units {a, b, c, ...}; therefore, if the theory can be expressed in terms of the sub-units, its range of applicability may increase.) As we see in the adaptationist/ bauplanist debate, sometimes rifts form in scientific communities when there is disagreement over the significance of the units of explanation. This is obvious in the point of contention among bauplanists and adaptationists, but the pattern is repeated in other areas of investigation as well. Ray Birdwhistell, a pioneer in the field of kinesics (roughly, "body language"), coined the term *kine* to denote an "atom" of movement. He was interested in categorizing and assigning meaning to various patterns of human movements, and thus he needed to devise a kinesic "language." His contention was that each kine was itself meaningful (1952), an assertion which some find unpalatable. The counter argument sounds much like the opposition to the strict adaptationist agenda:

> I find that this basic assumption is the most difficult one to accept. Perhaps scratching the nose is an indication of disagreement, but it may also be an indication of an itchy nose. This is where the real trouble in kinesics lies, in separating the significant from the insignificant gestures, the meaningful from the purely random, or from the carefully learned. (Fast 1970: 149)

In other words the issue is one of significance, just as in the adaptationist/bauplanist debate: What significance does a given phenomenon have in the context of a debate?

In approaching the question of how features of organisms are selected as the standards for determining normality, we can commit to a position different from the explicit views both of Gould and Lewontin (as sometime adaptationists) on the one hand and of Dennett (as hyper-adaptationist) on the other. What makes more sense is to allow that any interesting feature of an organism *might* be the result of adaptation or it might occur not as an adaptive response to a selective environment but merely as consistent with (not wholly determined by) other aspects of the organism's structure. But our original problem thus remains unsolved. We still lack a rigorous standard by which to atomize an organism for purposes of analysis, and consequently, we still lack an unchanging baseline for deciding questions about normality.

The overall problem here can be viewed as a variant of a long-standing debate between nominalism and other perspectives on what it means to categorize the objects of experience. To engage in the process of building a taxonomy one must accept a set of criteria as foundational. If the categories are to be more than the ephemeral visions of nominalists, one must interpret *foundational* here not merely as *what is necessary for the purposes of the moment*, but rather as *primitives* which in some sense are, if not essences of the organism, at least indicative of common structure (according to morphological taxonomy) or lineage (by the account of cladistic taxonomy). These foundational qualities will become the predicates in inductive arguments. This is the sense in which a notion of normality is at once dangerous and essential to evolution by natural selection.

To make sense of how kinds of organisms can evolve into other, discernible types, it might be claimed that we need to dump any commitment to essence. Thus Putnam (1990) argues that

> ...this Darwinian attitude, the attitude that says that the reality is the individual with all his uniqueness, his variation, opens the way for the idea that species slide into one another -- exactly what 'Aristotelians' thought was prohibited. If there is an eternal essence of Ape and an eternal essence of Human, how can one of those slide into the other? But once you say 'All there is is variation,' all there is is individuals in their variety, you have totally changed the picture. We might say, in this respect, Darwin was the most 'pragmatist' of scientists. (1990: 236)

In this case, to jettison the concept of essence is apparently not to abandon the concept of what constitutes the *normal* member of each species. But if such a concept is based on a rational standard (that is, if there are grounds for choosing the defining qualities of each taxon and the parameters within which they must fall), the concept is dangerously close to being essentialist if it is not in fact already so.

Evolution by natural selection makes statements about the relationships existing among species, not just among individuals or random groupings of individuals. Even if we acknowledge Putnam's point that Darwin was in some sense an anti-essentialist, natural selection as a theory is interesting because it talks about groups. Viewed at any given time, these groups do seem to have essences of some kind -- they must, else there would be no "glue" to hold certain individuals together under the same concept. Of course this does not necessarily mean the kind of essence associated with eternal species; it just means a quality or qualities of sufficient stability to make sense of the concept of natural kind. One can recognize the abstract problem of taxonomy in general (as 1993 Sober does) in much the same way as Quine (1960--the "Gavagai" phenomenon). Alas, this analysis of the problem does not provide a definitive answer as to how one can have a viable evolutionary theory of species without first countering the charge that species are themselves arbitrarily conceived. Clearly the central question here -- Are taxa necessarily arbitrary divisions of the continuum of life? -- is closely related to the question of whether there exists a standard for determining what is normal. When Gould asserts that bacteria are the most common sort of life on earth, it is obvious that we are presuming a reasonably fixed set of criteria for determining whether something is a bacterium, a set which can also tell us what is normal for a bacterium. But how did we choose these criteria? The pragmatic answer implied by Putnam is just one possibility.

A further difficulty for the concept of species results from the "stretching" of our concept of natural kind to accommodate new observations. The functionalist approach to categorization can stretch any scheme of classification to the breaking point. Even concepts which we would ordinarily consider absolutely firm and immutable are sometimes bent out of shape to suit the purposes of investigators. What could be more certain and unchangeable than an abstract concept of a number such as "two"? But recall the numerical analyst's joke from chapter three (2.(3)): "Two plus two equals five, for sufficiently large values of two." Anyone who has spent much time around numerical analysts knows that the round-off errors which accumulate during lengthy, complicated calculations can cause one to use names like "two" very loosely. An oft-heard phrase is "large value of" applied to a number such as 2, implying that the number has been augmented by round-off errors. Taxonomy includes similar stretches in which classification of organisms violates what non-specialists might well consider to be a *sine qua non* of a given kind of organism. Futuyma offers a diagram of "a fly in the family Nycteribiidae," which he goes on to describe as "wingless, eyeless" (1986: 119, 549). Surely the notion of a wingless "fly" is very nearly as hard for most of us to grasp as a kind of "two" which exceeds half of four.

A further difficulty emerges from the proposition that among some species there may be a genotypically determined capacity for heritable adaptation without extended cycles of reproduction. That phrase may require clarification. Motivated in part by the challenge of explaining how present-day biological diversity and complexity could have arisen since prokaryotic life emerged on earth perhaps 3.5 billion years ago, theorists have sought explanatory mechanisms which remain true to the Darwinian foundation of evolution by natural selection while simultaneously offering the possibility of "moving faster" than explanations relying only on selection of randomly occurring mutations as long-lasting adaptation (cf. Dennett 1995, 77 ff.). Other researchers served the same end as they pursued other agendas. Baldwin (1896), for instance, sought a means of reintroducing mind as an efficacious agent within Darwinian parameters. The essence of Baldwinian agency is a heritable genotypic capacity which allows individual organisms to change their phenotypic behavior. Thus phenotypes can sometimes occur independently of their associated genotypes in the sense that they are not uniquely determined by any given genotype. The effect remains squarely within the conceptual boundaries of Darwinian evolution by natural selection, however, because the capacity for the individual to alter its phenotype is determined by the inherited genotype. Thus the Baldwin Effect allows quick *en masse* convergence toward an enhancement -- quicker than if the species had to grope its way forward aided only by purely random variations among its individuals. All that is necessary, in Dennett's words, is the ability to "recognize" an advantageous variation. A question should immediately arise: How are we to define this "recognition" of advantageous variation apart from some concept of "memory"?

After all, an advantage, something better than something else, is a *comparative* notion. One cannot know an adaptation to be advantageous without simultaneously comparing the new state to previous ones. In short, there must be some memory of what went before (though of course we are using the term "memory" metaphorically). If the memory is really just a statistical tendency to reproduce and survive more effectively in one state rather than another, what we have would seem to be no special "effect" but rather simple ontogenetic adaptation.) What then is the *norm* for an organism possessing a Baldwinian capability to optimize its phenotype? If we answer that its macro structure is the norm, we may end up separating taxa which are very similar genotypically because of environmentally determined phenotypic differences. If we focus on genotype, we may group together individuals with markedly different phenotypes.

It is already obvious why one would claim that species are arbitrary groupings. Three methods of attacking this claim immediately present themselves. First, it might be claimed that the concept of homology provides a certain means of narrowing the field of candidate properties with respect to discussions of evolution. "...Traits like the body shape of sharks and dolphins are not used in classification. Biological classifications are defined instead by similarities that are not functionally necessary homologies as we are here using that term" (Ridley 1985: 8). The second counter to the charge that taxa are established arbitrarily is that taxonomy as a process is performed *functionally*. The groupings are indeed metaphysically arbitrary but are epistemologically meaningful. In other words, our reasons for paying attention to the groups are simultaneously the means of defining the groups. Thus by this account the allegation of arbitrariness is once removed but still alive and well. The third negative response to the charge of arbitrariness is what could be called a Darwinian variation of the functionalist answer: Our reasons for perceiving groupings as we do depend on purposes which are more or less universal. A strong form of this argument holds that universal grounds for perceiving organisms in groups are indeed instinctual. The foundations of taxonomy thus belong to *us* rather than to any *necessary* metaphysical reality, but our taxonomies reflect *a* reality and are not the function of mere whim. Pinker writes:

> In an important sense, there really are things and kinds of things and actions out there in the world, and our mind is designed to find them and to label them with words.

> That important sense is Darwin's. It's a jungle out there, and the organism designed to make successful predictions about what is going to happen next will leave behind more babies designed just like it....
>
> Moreover, it pays to give objects several labels in mentalese, designating different-sized categories "cottontail rabbit," "rabbit," "mammal," "animal," and "living thing." (1994: 154-155)

This seems a sensible reflection, but it casts discussion of Darwinian evolution in an odd light, assuming that such discussion depends on the identification of taxa. Suppose we define taxa as those groupings which it is *advantageous* to perceive. In other words, a selective environment gives an advantage to those who (1) perceive individuals as members of groups and (2) define the groups in certain ways. If this is the only basis of defining taxa, and if taxa are indeed necessary components in any defense of evolution by natural selection, then it may appear that any discussion of evolution is circular in a certain way: we can't make sense of evolution by natural selection as a model explaining why there are species without explaining what "species" means, which we do by saying that they are those things which evolution by natural selection drives us to perceive. (Pinker does not address the point.) This is where our previous reflections on recursive fitness can come into play. They tell us that it is not just the static form of some arguments which is significant, but their dynamic character. What looks to be a regrettable circularity in a static representation can prove to be a productive inquiry when perceived in its dynamic character. Much depends on the temporal element which we reviewed in our first discussion of recursion (chapter three). Our primary problem in this section -- how we establish normality by selecting qualities of organisms as their defining traits -- leads to a secondary one: How can we determine what is normal without reasoning in a circular way? The case we have just seen, in which the explanation of natural selection depends upon conceiving taxa which are themselves defined by appeal to natural selection, is just one example of the problem of circularity as it relates to concepts of normality.

### (2) Homology

Since the concept of homology figured importantly in the previous section, we may profit by considering the nature of homologies more closely now. Homologies[41] form a large part of the evidence for evolution in general and more specifically for establishing particular lineages. Homologous traits of organisms are structural similarities held in common by organisms which appear to be very different from one another. For instance, humans, moles, horses, birds, bats, pterodactyls, ichthyosaurs, porpoises, and labyrinthodont amphibians all have limbs -- arms, legs, wings, or fins, as the case may be -- composed of somewhat similar bones in somewhat similar arrangement: the humerus is the bone nearest the body, then come the radius, ulna, carpals (sometimes), metacarpals, and from one to five digits (Futuyma 1983: 47). Such shared substructures are taken as evidence for evolution because (it is argued) descent from a common ancestor is the best way of explaining underlying, partial similarities among organisms whose overall structures are somewhat dissimilar. Descent with modification encapsulates the evolutionist's general rhetorical program.

Of course theories relying on homologous characteristics must have some standard of similarity. In other words, the theories have to include criteria for deciding how similar two structures must be to justify inferring a common ancestor at a given point in the evolutionary tree. In addition, such theories must assume that descent from a common ancestor is the only or at least the best way of explaining the similarities ("best" being judged by some criterion such as parsimony). Let us take these issues -- standards of similarity and explanations of common design -- one at a time.

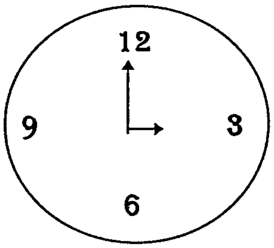
How similar is similar? That is the question which theories based on homologies must answer. The ulna and radius of a bat's wing are identifiable, or at least there are two bones between what we call the bat's humerus and the structure containing what we call the metacarpals. The same could be said of the arm of a human being. But the human's ulna doesn't look like the bat's in its shape or in its proportion to the radius. Are the two ulnae similar enough to be called a homology linking bats and humans to some distant common ancestor? The answer has to do with comparison -- how different non-homologous limbs of other species might look -- as well as with surrounding similarities. Forgetting for a moment that the carpals seem to be missing on the bat, and ignoring that the shape of the bat's ulna and its proportion relative to the radius differ significantly from what is found in the human, the similarities of the arm and the wing are striking: a somewhat similar humerus in each case, and five jointed metacarpals as well. Because of these similarities, the observer has some grounds to "fill in the blanks" by attributing the same names to structures which seem to be less overtly similar than the surrounding structures.

Here is a recent summary of "the homology criteria" (Haszprunar 1994: 135; my trans.):

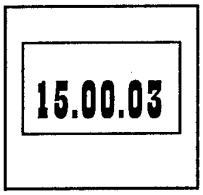
> The probability of homology increases (or the probability of analogy sinks) when:
>
> A) LOCATION/STRUCTURE:
> 1) the character* always takes the same relative position.
> 2) the character always has the same structure.
> 3) the structure of the character demonstrates high complexity.
> 4) the variations of the structure can be bridged by intermediate forms.
> 5) the character always demonstrates the same developmental pattern.
>
> B) CHARACTER DISTRIBUTION (PATTERN):
> 6) other characters show identical or hierarchical ... distribution (compatibility)
> 7) the character appears in all representatives of the group (constancy)
> 8) The character appears only in this group (exclusivity)
>
> C) FUNCTION AND ECOLOGY:
> 9) the distribution of characters is not correlated with a specific adaptation or a specific habitat (ecological criterion).
> 10) the character has little adaptive value (the so-called Darwin Principle).
> 11) the higher the number of basic (known) solutions for the function of the character is (the functional criterion).
>
> *By "character" is meant all the spatial and/or temporal patterns of an organism. Along with purely structural characters, physiological (e.g., warm-bloodedness), embryological ..., neurological ([measured by,]e.g., EEG, phonogram), ecological ... or ethological characteristics are also used as characters. The homology criteria apply for all types of characters.

The length of the list is an indication of the challenge posed by homologies. Taxonomy based on morphology must exploit common structure, but how can we be sure that two organisms which share a character are related? Notice that the list of questions applies to rising and falling probabilities; there is no chance of utter certainty. Attribution of homologous traits depends on judging what we might call structural contexts to be similar, and then asserting some degree of equality among the respective elements of the context. The overall similarity of the context leads to equating even elements which, viewed in isolation, seem rather more dissimilar than similar based on criteria such as size and shape. In the case of the ulnae of humans and bats, it is order or position within allegedly similar macro-structures which prompts the allegation of identity. (Here "identity" does not mean utter sameness, of course, but rather similarity sufficiently broad to warrant identifying the two bones with the same name.) It is immediately clear that there is a kind of circularity in this process of identification of dissimilar elements, which is not to say that agreement on homologies is always unwarranted. The process could be described in the following way: First, certain features -- boundary elements, we might call them -- are chosen from among the myriad characteristics which we can observe in a single species. A macro-structure such as a limb is demarcated in this way. Next corresponding features are found in another organism, one which, based on the totality of its observable features, appears to be different in structure than the first organism. On the basis of the structures which have already been identified as similar, a macro-structure is demarcated in the second organism. Within the two analogous macro-structures (a human arm and a bat wing, say) certain structures are singled out as identical by some criteria. The extremity of what we call the metacarpals might be one such criterion: these bones are furthest from the body in each case. Further likeness is to be found in the number of bones and the number of joints within each. Having identified the macro-structures, the observer then returns to appraising the similarity of the units within those boundaries.

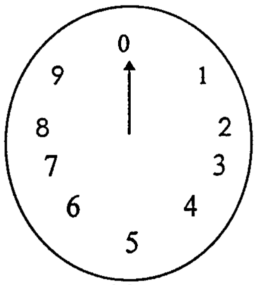
An allegorically similar case: Suppose that tomorrow someone finds Paley's watch lying in the sand. It's a late late-eighteenth-century model, but by some chance it is reasonably well preserved. And the coincidences aren't over yet! Next to the watch lies a Timex digital, manufactured earlier this year, and next to it is a modern analog altimeter (no doubt dropped out of a passing airplane). Here's how the find looks:



Paley's Watch



Modern Timex



Altimeter

A problem here is trying to decide in what sense any of the structures displayed might be homologous to one another. Are the similar shapes of the watch and altimeter sufficient to establish a homology? Tough to say. The criteria under A and B in Harzsprunar's list above would argue for homology, but the stanards under C complicate matters. What about the fact that there are digits around the edges? Again, we cannot be sure. Sadly, the two watches might well be thought less similar to one another than Paley's watch and the altimeter, at least if the observer pays attention only to a "snapshot" of structure rather than noting synchronization of movements (assuming the two watches are still functioning properly).

In short, composition, order, and shape are some of the criteria which we take as a warrant for using the same name to refer to a substructure of the bat wing and of the human arm. But at some point the set of similar features plummets to one -- order with respect to the rest of the macro-structure's elements in the case of the ulna, for instance. The process could be described more abstractly as follows: elements i and i' (the ulnae) are seen to belong to sets S and S' (the human arm and bat wing), respectively. Based on multi-faceted similarities (shape, micro-structure, order) among elements a, b,...,n and a', b',...,n', excluding i and i', S and S' are called equivalent in some sense. Because of a single similarity (order) between i and i' within these sets, i and i' are likewise declared equivalent. More briefly put, within somewhat equivalent sets corresponding elements are deemed equivalent.

### (3) Second problem: "Circular" normality

Count up all the mornings, count up all the sunrises, and compare: the relative sizes of numerator and denominator in this ratio are hopefully nearly enough equal to inform us that (1) It is *normal* for the sun to rise each morning. But isn't that analytic, since "morning" is itself defined by the occurrence of sunrise? Then perhaps we can say that (2) N*ormally* the sun rises once every twenty-four hours. This revised claim is liable to a similar objection, since our notion of "hour" is ultimately based on the solar period. Both statements (1) and (2) employing the concept of normality are true. Arguably the second statement is also informative (not analytic) for anyone who understands the concept of hour in terms unrelated to astronomy, as a railway station master may understand that the Boston and Maine train leaves Porter Square in Cambridge, Massachusetts, heading in the direction of Waltham once every hour. In this case, the statement that the sun rises once every twenty-four hours bears a meaning other than that the sun rises about once every time one twenty-fourth of the earth's period has elapsed twenty-four times.

It seems likely that such statements about what is normal -- circular with respect to the original meanings of their key terms but nevertheless informative when viewed from some epistemological perspectives -- can and do play a constructive role in the explanation of biological phenomena. Thus one could argue that "normal human heart" originally meant "the kind of heart that most people have" (with "kind" defined on a set of criteria as described above) but has since taken on a meaning that transcends the actual frequency of various kinds of hearts. If some future epidemic debilitates the cardiovascular systems of all but a few people around the globe (a variation of Wachbroit's counterfactual), we could then still talk of a certain kind of healthy heart as being "normal" even though it occurs *infrequently* in the species. Along these lines one could argue that although the discernment of natural kinds may originally have been motivated by the circumstances of a selective environment (as Pinker suggests), other, independent grounds for a given taxonomy may have emerged. Certainly one can imagine a survival advantage accruing to shepherds who realized that penning a ewe with a bull would be unlikely to augment the flock. Since those long-gone days, however, we have found ample justification (e.g., genetic similarity) for regarding sheep as a kind distinct from cattle. Below we will examine arguments that this non-circular understanding of taxonomy requires a commitment to a kind of reductionism. We will also examine the difficulty of establishing the normal -- in the sense of the limit to which organisms of a taxon tend -- as the product of a predictable cycle.

In so far as the *original* understanding of a norm defining a natural kind had to do with actual frequency, the concept is in some sense still rooted in statistics and is certainly not unique to biology among the sciences (contrary to Wachbroit's claim with regard to "biological normality," especially on p. 587). Is there any other sense in which one might acknowledge the counting of occurrences as the basis of "normality" and yet free the concept from quantitative definition? We will see in the next section.

### 3. F-normality (counting by comparing)

Supposing we have ignored or somehow avoided the two pitfalls of irrelevance and circularity, the concept of statistical normality may still require clarification.

Arguably the notions of counting and comparing can be understood in two ways. One sense of this process is clearly the notion of establishing and comparing precise *quantities*, as when one counts the number of organisms in which blood circulates and then the number of organisms with hearts, and finally compares the two numbers. But perhaps the process need not depend on exact numbers at all. Instead the observer might *relate* rather than *count*, in Frege's (1879) sense that "just as many as" can be established through a one-to-one relation without appeal to number. This relational sense is still quantitative in so far as "just as many as" has to do with numbers of things. However, specifying *definite* numbers is not a prerequisite. At a bicycle race one may see a pack of cyclists whiz past in an uncountable blur and yet still affirm with certainty that there were just as many riders as there were bicycles. To count n occurrences of some object can then be seen as a one-to-one mapping onto the set {1, ..., n}.

The point of mentioning Frege's definition of a cardinal number in terms of a one-to-one mapping is to imagine how a notion of statistical normality might be devised without appeal to counting in the most mundane sense. (Appealing to Frege certainly goes beyond Wachbroit's presentation of non-statistical normality, but *something* more needs to be said to justify a non-statistical conception of what is normal.) A very loose application of Frege's notion that we can compare quantities without counting provides this epistemological boost. What we cannot escape is counting itself in some sense. Horan is correct in her observation that "statistics is basically sophisticated counting (even for continuous variables, where the 'counting' is a matter of integration). Statistical properties are therefore *numerical* properties" (1994: 79). Still, if we conceive of number and therefore counting in a way that does not demand exact quantities, we may be able to grant some version of Wachbroit's claim that there is a viable sense of normality which is neither normative nor statistical (only we will read "dependent upon counting" instead of "statistical"). (Such a normality would, however, not be unique to biology and so should not be dubbed "biological normality.") The account of this normality would depend on a *mapping* of certain kinds of organisms with given properties onto some other set. For instance, over many generations hearts of a certain sort may be "mapped onto" (associated with) healthy human beings, although no definite numbers of humans or of hearts can be offered as evidence. It seems an exaggeration to say that this sort of mapping is not "statistical" -- it does after all rely on numerical comparison to the extent that one can say what sorts of association occur most often and which less frequently -- but such a mapping does indeed differ from a generalization made on the basis of carefully selecting relevant factors, counting them, and then inferring what is normal.

We can call the "mapping" kind of normality "F-normality" (acknowledging Frege's definition of cardinal number) as distinguished from the kind of "statistical normality" inferred from comparing exact frequencies of carefully abstracted properties. Based on this distinction one might guess that evolutionary theory depends largely, even entirely, on statistical normality. After all, the painstaking work of analyzing existing organisms, comparing their properties with those of organisms analyzable only through paleontological evidence, and finally adjusting taxonomies and fine-tuning cladograms would seem to require very careful selection of taxa-defining properties and then an exact counting to ensure that our taxonomies allow a rational comparison of structures relevant to speculation about evolution.

But perhaps this is not the case. Sometimes the reason why taxa are reclassified in the face of new paleontological evidence is that a new lineage is inferred. In other cases classification is debated because existing, observable properties are weighted differently by different camps of taxonomy. Are viruses alive? Do bacteria belong to the kingdom Monera or Protista? These are examples of debates over carefully abstracted and observed properties -- properties long seen as relevant to such discussions. But surely there are also ontologically real properties of organisms which are ignored in debates over classification. Genetic make-up is an obvious example of a feature of organisms which, though not taken into account at one time, now plays a crucial role in classification. An ongoing debate over the intraspecies division of humankind provides an example of how recent genetic investigations affect our perception of natural kinds. Pinker discusses the speculation of some linguists that a proto-language called SCAN may be the common ancestor of all Eurasian, American, and northern African languages. He notes that

> One interesting parallel [to this linguistic hypothesis] is that what most people think of as the Mongoloid or Oriental race on the basis of superficial facial features and skin coloring may have no biological reality. In Cavalli-Sforza's genetic family tree, northeast Asians such as Siberians, Japanese, and Koreans are more similar to Europeans than to southeast Asians such as Chinese and Thai. (1994: 257)

Pinker's phrase "biological reality" is particularly interesting in this context. Presumably scientists a few generations ago would certainly have considered facial features and skin coloring to be neither superficial nor the product of biological hallucinations. On the other hand, these traits *could* be taken as irrelevant to the agenda of delimiting natural kinds based on the research Pinker cites. Wasserstrom notes that by one account "a non-racist society would be one in which an individual's race was of no more significance in any of these three areas than is eye color today" (1977; in Hudlin's 1993: 31). Such a situation is possible in a moral sense; perhaps it would be possible in a scientific one as well. Or perhaps one could perceive some characteristic of organisms but attribute a much different worth to it than is currently done. It is reported that "Papa Doc" Duvalier, then president of Haiti, surprised a group of journalists by asserting that his island nation was inhabited almost exclusively by whites. One of the journalists, an American, pressed the ruler for clarification. Duvalier is said to have replied by asking a question of his own: "In the United States, is there a minimum percentage of 'black blood' which one must have in order to be considered black?" The journalist assured him that even "one drop of black blood" -- a metaphor meaning the presence of even one black person among the subject's ancestors (assuming the presence was known to one's community) -- was at one time sufficient to make the person "black" regardless of appearance. "We perceive our whites in the same way," Duvalier said, according to the story, and naturally he could claim QED at the same time: If the presence of any "white blood" suffices to make one white, then perhaps Haiti's population *is* overwhelmingly white. Of course it would be impossible to tell for sure how many citizens of any nation are of "pure" blood. It is not clear that a contemporary geneticist using the best of modern technology and know-how could say with certainty whether any of one's ancestors a few generations removed were light- or dark-skinned. Moreover, as Duvalier's quip makes clear, we do not have a clear and unequivocal standard for determining when the terms "white" and "black" properly apply.[42]

The wit of dictators aside, the advent of genetics has in large measure changed the way we think about natural kinds. But long before genetic explanations were available, invisible structure had been assumed to be the cause of observable phenomena relevant to taxonomy. We saw above that psychiatrist Oliver Sacks appeals to a hidden, "almost quantitative" difference in structure distinguishing the normal from disturbed patients. Ability to reproduce has been a common basis for determining membership in a given species, and this ability was naturally attibutable to *something* peculiar to the taxon. But for various reasons such abilities, whether or not they can be explanatorily reduced to genetic or even biochemical grounds, are weighted differently in different taxonomic schemes. In fact one expects to find such differences of opinion among cladistic and structure-based schools of taxonomy. Ereshefsky (1994) cites the example of two kinds of trees, long held to be different species, which nevertheless can reproduce with one another. One taxonomic camp asserts the two kinds are indeed different species based on their structural differences; the other school holds that the ability to reproduce suffices to demonstrate that the two "kinds" in fact constitute just one species. (See also Stanford 1995: 74 - 75, especially with respect to hybrids and syngameons.) Benditt reports a similar sort of controversy over the relationships among humans, gorillas, and chimpanzees following research conducted by Goodman, Miyamoto and Slightom. This research yielded genetic evidence suggesting that chimps are more closely related to humans than to gorillas. Or put another way, "...the simplest [branching] pattern [in a tree accounting for the genetic evidence] is one in which the gorilla first splits off from the progenitor of chimpanzees and human beings; later the chimpanzee and human lineages diverge to yield their modern equivalents" (Benditt 1988: 18). Prior to the publication of these findings, taxonomists had considered chimpanzees and gorillas to be each other's closest relatives based on their shared trait of knuckle-walking among other morphological similarities. Benditt concludes that

> It will undoubtedly be some time before the apparent conflict between morphological and molecular data is resolved. Meanwhile Goodman proposes that chimpanzees, gorillas and human beings be put in the same subfamily in the overall scheme of classification of the species. That would be a radical move, because they are now not only in different subfamilies but also in entirely different families: Hominidae for human beings and Pongidae for the apes (Benditt 18).

Yet another way of looking at this problem is provided by James Lennox in his explication of the concept of *genos* in Aristotle's biology (1987). Contrary to the view that the notion of *genos* is the primary taxonomic division in Aristotelian thought, Lennox argues (following Pellegrin) that "*eidos* (or *genos*) can refer to organisms at different levels of generality" (348) and later concludes that "it is clear that the term [*genos*] has no fixed taxonomic reference [in Aristotle's biology]" (349). By Lennox's account, Aristotle certainly uses *genos* as an organizational tool, but it is not a tool which allows one to construct a monolithic, "true" taxonomy which "carves nature at the joints." Instead, the Aristotelian biologist may use the concept to group organisms based on one or a set of qualitatively common features. Such a grouping depends upon purpose and perspective. This is not to claim that Aristotle had no concept of natural kinds as reflecting real differences among organisms, independent of the purposes of the observer. Lennox's exegesis does go to show, however, that the figure often held up as the archetype of taxonomic realism (by Putnam above, for instance) is actually very subtle in the way he employs the concept of division. In fact, his categories of classification themselves (e.g., *genos*, *eidos*) do not correspond to a single biological reality but are rather much more general in their potential applications.

What this all goes to show is that our concept of normality, based on *some* kind of counting of properties and organisms, must be done against a backdrop of other commitments and purposes. By some accounts of the process of induction in general, recognizing this backdrop forms the basis for the distinction between *confirmation* and *corroboration* on the one hand and *support* on the other. Thus Gillies (1988) claims that

> ...in science and everyday life, we use the notion of evidence (*e*) *confirming* or *corroborating* a hypothesis (*h*) or a prediction (*a*). We shall use the term 'confirmation' and 'corroboration' as synonyms, and write the degree of confirmation (or corroboration) of h given e as C (*h*, *e*). Strictly speaking, the evidence *e* will be in addition to some background knowledge *b*. So we ought really to write C (*h*, *e* & *b*). (1988: 181)

A bit later he continues by noting:

> Many confusions have arisen from conflating confirmation (or corroboration) on the one hand with *support* on the other. The difference is this. C (*h*, *e* & *b*) stands for the total confirmation given to *h* by both *e* and *b*. Degree of support S (*h*, *e*, *b*) is a 3-place function, and represents the contribution made to the total confirmation by the individual item of evidence *e* against a background *b*. (ibid.)

### (1) F-Normality and Fitness

If the concept of probability is not to take center stage in the definition of fitness (as it does in Mills and Beatty's propensity interpretation), then we must draw on some other reservoir of data to make the sorts of comparisons in which fitness usually figures. There seem to be two such general types: first, there are comparisons between two specific individuals, and then there are comparisons between one individual and a group of individuals from the same taxon and in the same environment. One might expand the list with comparisons between the same individual's respective fitnesses in different environments, or perhaps by adding comparisons between the respective fitnesses of two individuals in different environments (Is a polar bear in an area north of the Arctic Circle fitter than a rattle snake in the Mohave Desert?), but the important thing to realize is that the comparisons which evolutionary biologists study can be distilled into pairs, and the pairs focus on individual and group fitness. What I want to argue below is that we can leave probability and related terms -- tendency, propensity, and so forth -- out of the definition of fitness and out of claims about fitness, provided that we view fitness as belonging to types and not to individuals *per se*. Part of this argument has already been made, especially in chapter eight, 1.(1); the goal of raising the issue again here is to show how we might investigate normality without explicit appeal to probability based on exact counting.

### (2) Background versus environment

"Background" in Gillies' sense should not be conflated with a factor such as environment in the simplest interpretation of that term. Certainly environment is crucial to an understanding of how traits are selected (Brandon 1990). In fact, environment may be the key factor in explaining how one can describe long-isolated lineages as fit and how we can make sense of fitness in the case of the conflict recently alleged to exist between the human mother and fetus (in which a fetus's interests, metaphorically speaking, are somewhat contradictory to the mother's in that both must share the same resources; see below). One problem with viewing fitness as a propensity is that raw numbers of species members may not indicate a likelihood of survival in situations where isolation is a key aspect of the environment. On the Galapagos Islands, for instance, the introduction of rats on visiting ships posed a particular peril to long-isolated species of birds. From time immemorial these birds had nested on the ground with impunity because of the lack of predators. In fact many species or even cultures seem almost more vulnerable when they are particularly populous precisely because their apparent adaptedness is a two-edged sword. One thinks of the kiwi as well. A cultural metaphor may be instructive at this point. Military historian John Keegan notes:

> It was [the classical Greeks] who, in the fifth century BC, cut loose from the constraints of the primitive style, with its respect above all for ritual in war, and adopted the practice of the face-to-face battle to the death. This departure, confined initially to warfare among the Greeks themselves, was deeply shocking to those outside the Greek world who were first exposed to it. The story of Alexander the Great's encounter with Persia, an empire whose style of warmaking contained elements both of primitive ritual and of the horse warrior's evasiveness, is both real history, as narrated by Arrian, and a paradigm of cultural difference. (Keegan 1984: 389)

The scenario which unfolded is predictable: Alexander wreaks large-scale havoc on the Persians, an apparently mighty and thriving culture, because the empire could not adjust to the radically new threat embodied in non-ritual, almost Clausewitzian total warfare. Similarly, species which are numerous (i.e., enjoy huge reproductive success on average) may thrive because they have found means of isolating themselves from predators. Does this mean they are *likely* to survive except in the case of large-scale natural catastrophe? By the classical definition of fitness as a function of immediate reproductive success, yes. It all depends on how the selective environment is defined and where its borders are placed. But perhaps there is a workable alternative, one that attributes fitness "brownie points" on the basis of equilibrium within an environment of constant and varied predation.

Competition among members of a taxon, even of those closely related to one another, is frequently observed -- as when siblings battle for feeding position at their mother's breasts or when fathers kill their own offspring. There is currently also speculation about self-interest at some level causing *in utero* warfare between mother and fetus, with the fetus employing hormonal methods to grow as large as possible as quickly as possible in order to have a better chance of post-natal survival. The mother thus sometimes suffers hypertension, even endoclemia, and must survive the passage of a baby-saurus down the birth canal. (Of course "interest" is here not to be taken as a *conscious* longing for a certain state but must rather be understood metaphorically.) This is the scenario envisioned by Haig as reported by Horgan (1995). To describe the relationship between the two "players," one might attempt to construct a fitness/advantage matrix such as Sober does as he relates foot speed and resistance to disease (1993: 58-59). In the case of the mother, the point of such a matrix would be to quantify her interest in maintaining the fetus in such a state that she does not suffer from severe hypertension and further to ensure that the fetus remains small enough to pass with relative ease down the birth canal. In the case of the fetus, the matrix would place a numerical value on the fetus's interest in becoming as large as possible in utero. But what values does one assign to the variables in the matrix? On the mother's part, there must be an interest in the live, even healthy birth of the fetus. Or perhaps that statement is not quite correct, since some mothers may bear ambivalence or even antipathy toward the fetus. In any case there is (to speak anthropopathically) an interest on the part of *something* self-interested -- a selfish gene or egocentric species, perhaps -- which the mother at least *represents*. On the other side, there is the fetus and its personal (metaphorical) interest in survival as an individual (or as a collection of selfish genes or as the member of a species). Clearly the newborn baby which the fetus will shortly become will need the mother to survive under some scenarios (in cases where no surrogate source for nursing will exist *post partum*, for instance). In fact, although arguably the newborn could become immediately independent of the mother in a scenario where another benevolent, lactating, well-nourished woman is present, it seems clear that having a living, healthy mother after birth is normally advantageous. If this is true, then no *in utero* conflict between mother and fetus could be considered a zero-sum game. If maternal hormones should keep a given fetus relatively small, for instance, it has gained something (an increased probability of survival because its mother will probably experience an easier birth than if it had grown larger) even though it has failed to achieve its maximum potential size.

Given this state of affairs, how can one quantify the interests involved and thereby test a hypothetical propensity? There can be no *a priori* determination even of a numerical inequality except in the case of the mother. That is, we cannot say that smaller size at birth is *more* disadvantageous than otherwise from the point of view of the fetus, the species, or the genes, without specifying a host of background conditions. We can only say that the mother benefits (is more likely to survive for a longer time) when the fetus is small. (This discounts any psychological distress a mother may experience over an abnormally small baby; it also ignores the possibility that rearing an underweight newborn may be much more difficult for her than caring for an infant of normal weight.) If researchers who posit a sort of *in utero* war are correct, the progress of conflict is governed either by a break-even point or else by pure chance ("pure" in the sense that there is no convergence toward a "truce" mediating the opposing interests). Considering the number of spontaneous abortions, still births, and fatalities among mothers during birth, it is at least conceivable that mother-fetus hostilities are analogous to a Hobbesian state of nature, of war of "every man against every man," where no biochemical social contract applies. The equilibrium which is established in the course of this ongoing conflict would then be observable only in the range of outcomes.

Quantification of advantage would still be problematic, however, since there will always be doubt as to the overall advantage or disadvantage of any conceivable outcome -- for instance, of a birth in which the infant is severely underweight and the mother experiences an easy delivery because of the infant's small size. The baby is more likely to die than it would be had its weight been greater; on the other hand, the mother is more likely to live, and more likely to conceive again quickly, in the event of a very easy birth. To quantify the advantage or disadvantage of a small infant from the perspective of, say, the species or the mother's genes, one would need to evaluate factors which are both complex and highly dependent on the environment. Presumably the mother's age and number of past children would bear on how many children she would be likely to have in the future. Measured against the probable reproductive capacity of the existing infant, one might be able to conclude in some cases, for instance, that the mother's survival is less crucial to the species' or her genes' survival than the infant's. There would then be a tendency to select for individuals in whom the fetus would win the *in utero* fight, whatever winning would mean in such a context. The only way to define what is advantageous would be by appeal to the statistics within a given environmental context.

For our purposes, one significance of Gillies' sense of background is to emphasize the difference between the evidence which we think supports our inductive speculations -- that a certain organism must be related to other organisms discovered by examination of paleontological evidence, for instance -- and the circumstances (background) which make that evidence possible. Presumably our account of the evolution of organisms would be significantly different if humans were, as a rule, blind. Another way of encapsulating this dependence of concept on background is provided in Kitcher's discussion of how one should understand the terms of a given theory: "I suggest that an expression-type used by a scientific community is associated with a set of events such that productions of tokens of that type by members of the community are normally initiated by an event in the associated set" (1978: 540). Kitcher's primary aim here is to defeat the relativism which he attributes to Kuhn and Feyerabend, but his analysis also reveals the ways in which the terms of a scientific theory depend upon "initiating events." For our purposes it is important to recognize that there is no clear-cut standard determining how these initiating events, the ones which give birth to discernment of natural kinds, occur. We have just seen functional standards, adaptive standards, and apparently arbitrary standards.

The ultimate point of the discussion in this section is that as the model scenarios on which our conclusions about fitness become increasingly complex, it becomes ever more difficult to determine what constitutes something like background (e.g., in Gillies sense) as opposed to environment. To whatever degree a propensity interpretation of fitness may help us avoid paradoxes stemming from random agencies, it does so only to the degree that the environment is held not just constant but also constantly *bounded*. But this leads to paradox of a different kind. Suppose we took a "snapshot" of a species of bird on the Galapagos islands in the 14th century and another 600 years later. We might judge that particular species to be extremely fit; then, rather suddenly (by the standard of evolutionarily significant time frames), the species is decimated even though its range has not changed a whit. We can say that its selective environment has changed over the course of six centuries, that is, the birds' fitness as a measure of reproductive rates observed at various intervals prior to the influx of new predators would be much different than their fitness after that event. But on the other hand, it would be logically possible to understand fitness as an average reckoned over a longer period and consistent with a more broadly conceived environment. In this way the bird species' fitness could be judged constant from its arrival on the islands (or the formation of the islands) through the present time. Now if the identification of time frames and environmental boundaries is arbitrary in this logical sense -- if there are no privileged criteria for establishing geographical and temporal parameters apart from the purposes of a given researcher -- then a propensity interpretation seems also to be somewhat arbitrary in so far as it relies on averages within arbitrary bounds. To repeat what has become an old refrain for us, a dynamic means of reckoning fitness (e.g., a recursive one) seems a preferable to static means in such an indefinite context of observation. By employing a recursive means of fitness which emphasizes the relationship of a value to its *immediately* previous state (as argument of a recursive function), our conception remains fluid and versatile. A recursive "step" can be as big or small as we choose for it to be, and that goes for any axis of measurement, including the geographical and temporal.

## 4. Existence and Stability

We can understand Darwin's agenda in the *Origin* as the attempt to build a model which could account for all the extant data on organic variation, including observations of existing species and of the fossil record. Part of this program required him to explain how and why species change within their respective environments. Without such an explanation one could not say why some species had apparently gone extinct nor account for variations among related organisms. Two key facets of organisms -- longevity and reproduction -- vary from individual to individual within a given species. Some individuals live longer and experience greater reproductive success than their peers. These phenomena of greater longevity and production of more descendants can result from chance, but they can also be explained by observing that some organisms do better in a given environment because of the way they are constructed and the way they behave. Now if these morphological and behavioral traits are heritable, then it stands to reason that certain lineages will thrive while other will tend to die out.

It may prove helpful to restate the foundation of Darwinian evolution in the following way. There are two major components in a species' history. First, there is the matter of its existence: at any given moment in time, a species has either ceased to exist or else it has at least one surviving member. Within this range -- from existence to non-existence -- there are myriad possibilities. A species or any other taxon can barely survive or can positively flourish, but such descriptions are subjective. The only elements of the spectrum which are objective are survival, extinction, and numerical counts of organisms. In other words, at a given moment the fact of the matter is that either there are living organisms of a certain type or not; in the first case, there are precisely X organisms of the type alive at a specific point in time (whether anyone knows it or not). But it is not necessarily the case that when a species is populous it is flourishing, nor that meager numbers mean imminent extinction. A colony of bacteria might number millions of individuals but still be threatened. If this colony were the only one of its kind, the species could be on the brink of extinction, depending on the conditions of the colony's environment.

The second major component of a species' history has to do with the question of what could be called its *stability*. The individual organisms which make up a species are just that -- individuals. Presumably no two organisms are precisely alike (a generalization which goes even for "identical" twins). But each organism of a taxon is sufficiently like the others to warrant membership in the same species. As each generation dies out, so do certain of the structural and behavioral differences disappear; with each new generation comes a new set of differences. Some of the differences may be adaptations, that is, aspects of morphology or behavior which environmental pressures have made increasingly more likely to appear among the members of the species, until at last what was a difference becomes the norm. When that difference is dramatic enough speciation has occurred.

Here the specific meanings of the nebulous terms "difference" and "dramatic" depend on the operative species concept of the observer; perhaps the classic definition is Mayr's concept of a reproductively isolated population of interbreeding organisms (1940: 254; cf. Aiello and Dean 1990: 3). That sounds clear-cut, but Mayr's conception permits the interbreeding population to exist "potentially" as well as "actually." The tension here becomes obvious if we ask ourselves, for instance, whether species which do not breed in the wild but can do so in the laboratory are really different species at all (q.v. Ridley 1985: 107 ff.). The component of a species' history which has to do with stability, then, can be represented as a spectrum ranging from what could be called *stasis*, in which no adaptation from any source (e.g., random walk, natural selection) occurs, to *speciation*, in which such dramatic, consistently repeated and widespread differences appear that an observer concludes a new species has emerged based on some taxonomic principle.

Two facets of the stability component warrant special emphasis. First, stasis is a relative term: presumably there are always some differences among the various generations of a species, but when these differences fall within a certain range determined by the particular species concept employed, it is fair to say that the species is in stasis from the perspective of evolution. Second, the judgment as to whether speciation has occurred is made only in hindsight. If two-headed pigs are the norm rather than the exception in a given generation, one must wait to see whether the characteristic appears consistently in succeeding generations before announcing the emergence of a new species of pig characterized by having two heads. Perhaps a genome mapping project will someday make it possible for researchers to know which variations are probably going to be repeated in future generations. Using a predetermined species concept, one which specifies the exact range within which each measurable characteristic of an organism must fall in order to belong to a given species, one could then announce on the spot that speciation has occurred or not. But for now we must wait an indeterminate period to see how ubiquitous a quality such as two-headedness becomes. Looking back in time, we can wield our species concepts to announce either that speciation has occurred or merely that differences among individuals or infrequent mutations have arisen within the acceptable boundaries of a certain species concept. It is clear that much depends on what range of possibilities a given species concept considers normal. Before turning to a more in-depth consideration of the species concepts and normality, however, we should look more closely at the two components of a species' history one at a time.

The lowest extreme of the existential component of a species is *extinction*, defined as the state in which a species has proved to be so unsuited to its environment that all of its members have died. But the extinction of a species can take place in two different circumstances. The species can die childless, as it were, having been the ancestor of no species which continues to exist. Alternatively, a species can become extinct after or while becoming a "parent" species. In this latter case, the ancestor species is extinct in a much different sense than in the first (see diagram below). But if genetic compatibility (evidenced by the ability to interbreed) and geographical range constitute the *sine qua non* which defines the members of a species (consistent with

Mayr's definition), then extinction would be much harder to define. Is the lineage of the Neandertals wholly extinct or were they a parent species of modern *Homo sapiens sapiens*? We don't know (Gore 1996). Thus theories of speciation normally include a morphological component, however small, except among the most extreme cladists (who as a school emphasize lineage over morphology). Given a morphological criterion, we know what extinction means: there are no more living organisms who can interbreed with other living organisms *and* who have certain morphological features.

Neither a specific term nor a concrete concept can be associated with the other end of the existential spectrum. What would "complete success" within an environment mean? Organisms can neither live nor multiply indefinitely; overpopulation is a sure ticket to decline. In other words, a group of organisms can be wholly unfit for their environment. As of a certain point in time -- 1993, say -- *T. Rex* had ceased to exist except in Crichton and Koepp's screenplay for a Steven Spielberg fantasy film (1993). But a species cannot be *optimally* suited to its environment. *T. Rex* and superstar movie directors such as Spielberg *thrive*, but they won't live forever nor will they reproduce indefinitely. Regardless of how well adapted a group of organisms is with respect to a given environment, some of the individuals of that taxon will be killed by environmental phenomena -- predation and climatic changes, for instance -- and some will fail to reproduce. (It is not clear that all individual organisms must "die," but those which are not predestined for mortality seem not to be individual except in an odd sense; q.v. Margulis 1994.)

Thus it can be said that fitness as an explanatory concept has a stronger negative than positive connotation in that the negative end of the first fitness "spectrum" (extinction-survival-"thriving") is well-defined while the positive pole is not. The concept opposed to extinction is simply *survival*, but since survival is not associated with a definite superlative or optimum in the way extinction is (*totally* dead), the most we can say is that there is a weaker concept of survival (surviving by the skin of the teeth or some such indefinite phrase expresses the matter as well as any other) and a stronger one (thriving, we could say, or perhaps an observer might assert that he sees no imminent danger of the species becoming extinct). The most accurate way to characterize the concept of survival as it applies to this fitness spectrum may be to say that it functions as a shorthand for "not-extinct."

The second major component of fitness is intended to explain another spectrum of phenomena besides the one bounded by extinction and survival, namely, certain wide-spread, long-term morphological and behavioral changes in organisms. Such changes are commonly called adaptations.[43] The "low end" of this spectrum is stasis -- no appreciable changes among members of an interbreeding population. At the other end is speciation, that is, adaptations so numerous or so extreme among the organisms of a species that it "splits," resulting in two populations which do not interbreed (or cannot, depending on the species concept). The two species may have the capacity to interbreed (be genetically capable of producing offspring), but morphology and reproductive behavior in the wild tell us that they are different species.

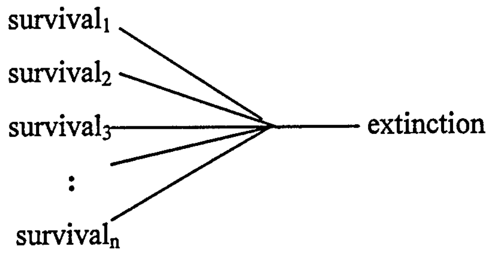
To summarize, then, the concept of species is associated with two spectra, from survival to extinction, and from stasis to speciation:

($F_1$) survival ---------------------------------- extinction

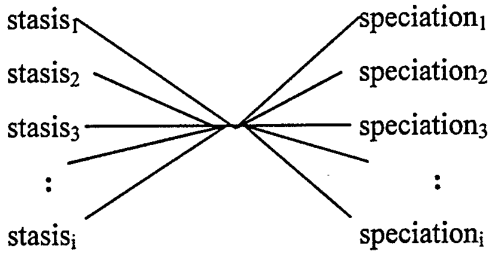
($F_2$) stasis ------------------------------------ speciation

As noted above, the right-hand pole of spectrum $F_1$ seems unambiguous. We know what extinction means: no survivors. The term "survival" is less well defined, since it is not clear what it means in the extreme, that is, as the opposite of extinction. We might reasonably associate extinction with the number zero. But survival could as well be associated with one individual as with one breeding pair as with millions of individuals. The poles of spectrum $F_2$ depend upon our species concept. By stasis we mean that a group of organisms remains *nearly enough* similar across generations so that we can say the taxon has not evolved into another. Speciation is the complementary concept, by which we describe a process of evolution dramatic enough to justify saying that a new species has evolved (in accordance with some concept of "species"). Accordingly, we might refine $F_1$ and $F_2$ as follows:

($F_{1.1}$ )



($F_{2.1}$ )



The pole "speciation" is ill-defined for another reason as well, namely, that the concept of species is itself ill-defined and is in fact the subject of ongoing, lively debate among evolutionary biologists. There is also a rhetorical advantage to be had from treating speciation as a nebulous concept. Futuyma, an evolutionist and defender of the materialistic faith against creationist claims, emphasizes again and again that taxonomic divisions are arbitrary. In the context of his book, *Science on Trial* (1983), this admission of arbitrariness is much more than an acknowledgment that species concepts are functional -- that is, that they depend on a researcher's purposes. (For instance, tracing genetic lineages of existing species may require a cladistic concept while attempting to generalize from sparse fossil evidence may require a species concept based on morphological comparisons.) What Futuyma means by arbitrary in this context has to do with the defense against creationist and saltationist claims. [44] He asserts that if neither the fossil record nor laboratory breeding experiments show gradual evolution from one species to another (a debatable claim given the fact that plant breeders have witnessed the emergence of tetraploid offspring from diploid parents). Futuyma implies that since species are arbitrary delineations of a continuous spectrum, the (non-saltationist) evolutionist need only show "significant" variation in the laboratory or the fossil record to prove that evolution is the best explanation of all extant data.

If the aim of the argument is to defend evolution by natural selection as an explanation of observable variation, then a circularity is obvious here. If we define evolution as the creation of variety, and define the latter term as nothing more than the result of variation without stipulating the extent of variation which "counts," then it is easy to find examples in which evolution is observable. In a short span of time scientists can observe significant random or environmentally-caused variation in fast-reproducing species such as fruit flies. By this account, in other words, one could redraw species boundaries in such a way that the fossil record and laboratory experiments, and perhaps even animal breeding in agriculture, would show speciation, and hence evolution. (Actually, one detects a degree of ambivalence on Futuyma's part throughout the book. On the one hand he makes the case just described, while also taking pains to suggest that leaps between existing species would not necessarily appear in the *known* fossil record for reasons of geological and scientific coincidence, and that laboratory experiments have had insufficient time to show variations large enough to represent actual speciation according to common demarcations of species.)

It is clear that while a wholly arbitrary species concept is helpful to those who argue against creationist and saltationist claims, the same arbitrariness could have social and political consequences which many would find undesirable. For example, it was long held that what we continue to call "races" of human beings were in fact not just varieties within a common species but were rather different species which developed from separate ancestors. Even if we omit the clause about distinct ancestry, the allegation that the races are separate species remains inflammatory. This is because "species" is a loaded term, implying *extreme* difference in structure, behavior, and potential (Mayr 1963).

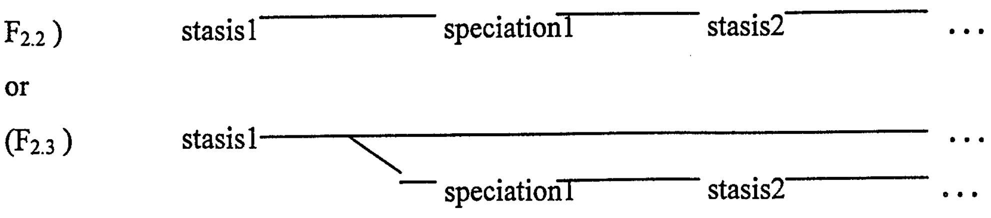
Dennett's *Darwin's Dangerous Idea* is a good indicator of how virulent the debate over gradualism versus saltationism can be. There Dennett the gradualist and hyper-materialist decries Stephen Jay Gould's support of the theory of "punctuated equilibrium" as an attack on the thoroughgoing materialism which Dennett thinks is a hallmark of Darwinian evolutionary theory (1995: 282 - 299). Paleontologist Gould (together with Eldredge 1977, 1993; see also Gould1980: esp. ch. 18) had sketched his dissatisfaction with thoroughgoing gradualist accounts of evolution. In place of complete gradualism, he posits the existence of macro-mutations -- sometimes corresponding to morphological leaps -- which ostensibly fulfill two explanatory purposes. First, they allow one to view the staccato fossil record without giving up the basic mechanistic rather than teleological mode of explanation which is at the center of the theory of evolution. Creationists of various ilks had insisted (and continue to assert to this day) that because the fossil record is not continuous, because it is in fact highly discontinuous, some sort of creative agent must have acted to bring forth various organisms *de novo*. But assuming that macro-mutations can arise suddenly and mechanistically from existing species, the fossil record remains evidence for evolution rather than for saltationistic creationism.

The second role played by the macromutation account again has to do with what had been perceived as a weakness in the theory of evolution given available evidence, but this time the leap between structures *within* a species was the issue. The basic problem was to explain how very complex structures -- eyes, for instance -- can develop gradually when the constituent parts of these structures would seem to offer no selective advantage. If a prerequisite of gradualism is that complex structures must have emerged slowly, then presumably this slow evolution is to be characterized by the assembly of individual parts, each of which offered a selective advantage or else was a "rider" on an advantageous feature. Above (pp. 64 ff.) we used the term "advantage clause" to refer to the requirement that every stage leading to a complex structure must be at least as advantageous as competitor structures. The explanatory task of dissecting a complex structure and then attributing a selective advantage to every individual constituent part as well as to intermediate combinations of parts has proved too great a challenge for some evolutionists. Gould is one such. Perhaps through a sort of scientist's instinct, these observers perceive a line between wishful thinking on the one hand and creative explanation on the other. At some point, these theoreticians feel, that line is crossed, and instructive science degenerates into misleading fantasy.

Enter the theory of punctuated equilibrium. By this account, complex structures such as eyes can emerge suddenly, complete or at least more nearly so than the gradualists allow. Under this scenario there is no need to explain the utility (i.e., the fitness advantage) of the parts of complex structures, although one might wish to examine these constituents to see if they *could* conform to a gradualistic model. It is enough to see that the complex structure itself confers an advantage on those organisms possessing it. Thus the persistence of the structure is explained in the usual way, while its origin is explained not quite as a creationist would account for it but also not completely in accordance with a gradual sort of mechanistic evolution.

Dennett asserts that gradual evolution by natural selection is sufficient to explain the extant data. (Dawkins makes essentially the same argument, emphasizing that gradual evolution and the available time suffice to explain the development of complex structures such as eyes -- 1995: 87 - 108) Phillip Johnson, Berkeley law professor and author of an analysis of Darwinian evolution from the standpoint of how the theory would hold up in an American court of law given rules of evidence and other conventions (1991: "Is Darwin Obsolete"), offers a purportedly objective analysis of the evolutionists' internecine strife over the issue. (I call this analysis "purportedly objective" because Johnson claims allegiance to no particular camp and claims to offer a purely logical evaluation of evolutionists' claims, albeit as a non-scientist.) He concludes that the evidence for evolution is simply insufficient. His position amounts to a rejection not just of the same arguments Dennett and Dawkins would later make, but also of Gould and Eldredge's argument for punctuated equilibrium (1971, 1993).[45]

The two fitness spectra above also seem to differ in respect of what we might call their reversibility. $F_1$ is presumably not reversible (though I see no way to prove that it is so, since the reversibility of certain physical processes may be merely improbable rather than impossible, q.v. Feynman 1965: esp. 108 - 126). A species' return from extinction in *precisely* its prior form seems at least so highly unlikely that it can be ruled out for practical purposes. Thus there is a directionality to the spectrum represented by $F_1$ and $F_{1.1}$. The case of $F_2$ and $F_{2.1}$ is less clear-cut. The path between a particular stasis and a certain instance of speciation is presumably directional, but in order for a group of "new" organisms (i.e., ones differing significantly from the original species) to be recognized as a species, they must also find a stasis, no matter how short-lived. In that sense, generations of evolving organisms travel back and forth, between stasis *in general* and speciation *in general*:



$F_{2.2}$ represents the case in which the entire original species evolves into a different species. $F_{2.3}$ shows the presumably more usual case in which the original species continues to exist for a time in parallel with the new species which emerged from it. From an epistemological point of view, speciation is partly a function of the following stasis. If a few organisms change their structure or behavior in what our current species concept tells us is a radical way, we must nonetheless wait to see if the new breed can indeed reproduce and survive before we say that speciation has occurred. In attempting to piece together the historical record of evolution, stasis is again epistemologically prior to speciation. We infer that speciation has occurred in order to explain the appearance of a stable species of a type which did not previously exist. Thus in $F_2$, $F_{2.1}$, $F_{2.2}$ and $F_{2.3}$, speciation can fall out entirely. It is understood that speciation must have occurred if we simply draw a line showing that one stasis emerges from another stasis. Of course the two only have meaning when compared with one another: neither one in isolation is sufficient to tell us that a phenomenon involving fitness has occurred.

($F_{1.2}$) extinction ———————— not-extinction<br>
($F_{1.3}$) extinction ———————— not-extinction$_1$ ($NE_1$)...$NE_2$...$NE_3$<br>
($F_{2.4}$) stasis$_1$ ———————— stasis$_2$: f (stasis$_1$)

Speaking of one stasis as *a function of* an earlier stasis may lead us to consider in what sense fitness itself, in so far as the concept figures in $F_1$, $F_2$, and their variants, is associated with a function. This question is of course a different sense of the term "function," but it is surely worth asking. Is it the role of fitness to move surviving organisms toward extinction? Or does fitness move organisms in the other direction, toward a greater harmony with their respective environments which in fact makes survival more likely? After all, organisms can move from one non-extinct state, $NE_i$, to another, $NE_n$ (with $n > i$) -- that is, toward a more flourishing state of existence as well as in the opposite direction, toward extinction.

Let us define a "ratchet function" as having two qualities: first, the function itself moves in only one direction; secondly, it prevents other aspects of the entities on which it is operative from moving backward, that is, to an earlier state. Then apparently a ratchet function *can* be operative in the first component of a species' history, that is, in the spectrum from extinction to survival. Once extinction has occurred, the given species does not reemerge in precisely its previous form. But again the issue of normality arises. What if a species -- of bacteria, say -- becomes extinct, but then a million years later, another, very similar form emerges due at least in part to natural selection? How similar need "very similar" under a given species concept be? It is at least conceivable that a species could reemerge after having been extinct, just as Feynman notes it is conceivable that all of the molecules of blue dye would, through random Brownian motion, separate themselves from the water molecules with which they were at one time mixed (Feynman 1965: 111 ff.).

The significance of the foregoing discussion for our overall interest in recursive fitness is signaled by the definition of one period of stasis in terms of the preceding period ($F_{2.4}$ above). We saw in the previous section that recursion is a good method for handling geographical and temporal boundaries of a concept of selective environment when those parameters can be set virtually anywhere or can even be transient from a purely logical point of view. Now if fitness describes a dynamic between organisms and their environment, we may well ask if there is a logical underdeterminacy of the boundaries between taxa such as species corresponding to the indeterminacy of the selective environment's defining parameters. In this section we have seen that the answer is positive -- we must view species concepts as at least somewhat arbitrary in the sense that empirical data neither wholly determine nor perhaps even privilege a given species concept compared to competing schemata. When we seek to investigate measures such as fitness which are sometimes associated with species concepts (as in judgments such as "That dark-colored species of moth is better adapted to the environment than the light-colored species over there"), we should therefore choose a definition of fitness which need not apply to any specific grouping of organisms associated with a particular period of stasis. Because a recursive function proceeds by defining a present state by *a* previous state of existence, we can choose the interval between the two states to be huge or vanishingly small or anything in between, depending on the purposes at hand. Again, the goal is to find a wholly content-*less* (formal) understanding of fitness for the sake of achieving maximum generality. Recursive fitness seems to provide such an understanding.

# Chapter Ten: Units of Fitness and Species Concepts

At the risk of being somewhat repetitive, this chapter attempts to drive home the notion that recursive fitness (and a recursive view of the hypothetico-deductive method in general) provides a way of avoiding if not resolving certain debates associated with the unit of selection and species concepts. I hope that a couple of new approaches, some odd terminology and a few examples will make the discussion worthwhile even where we have already encountered the conclusions about how recursive fitness can be a productive concept.

## 1. Units of selection, replication, fitness

The "unit of selection" is the topic of a long-standing debate which can be encapsulated in a single question: What is it that struggles to reproduce itself against hostile forces? We can think of some "level" of life as being analogous to a body in motion under the assumptions of classical mechanics: the body, once in motion, continues to move in a straight line until some force or combination of forces -- an obstacle, for instance, or simply ongoing friction -- perturbs, or hinders and eventually stops the motion. (What "level" means in this context will become clear momentarily.) The unit of selection debate seeks to find what it is that similarly progresses and can be analogously perturbed, or hindered and stopped, only in this case the change in question is self-replication rather than motion. Like the motion of a body, self-replication takes place (albeit at indeterminate intervals) until it is slowed and in the end stopped by other forces. Theoretically an organism might replicate itself *ad infinitum* unless interrupted by a counteracting force, just as a body moves in the same direction with the same velocity unless influenced by another impetus. Or we might say that replication as motion is perturbed (rather than slowed or stopped) if one organism evolves into another type of organism.

The phrase "body in motion" itself answers the question, What is the substrate of change? But in the case of replication, more than one response is possible. The classic view is that the individual is the substrate of evolutionary change in two senses: first, it is that thing upon which hindering or perturbing forces can act; second, it is that which reproduces itself. However, it is clear that other levels of life can fulfill at least this second role, self-replication. The "selfish gene" -- a term popularized by Dawkins' 1976 book but anticipated earlier by Williams (1966) -- exemplifies a level of life which arguably can self-replicate and which can therefore function as an alternative to the individual in this regard. But can the gene (or any other level) function in the first sense which a substrate of change must fulfill if it is to be seen as the unit of selection, namely that upon which hindering or perturbing forces can act? Before attempting to answer this question, we should consider selective forces themselves more carefully.

The hindering or perturbing forces relevant to the unit of selection debate can obviously be *external*: environmental factors such as heat and cold, flood and drought, light and darkness, can have positive or deleterious effects on some "level" of life's ability to reproduce itself. It is less clear whether hindering or perturbing forces can be *internal*. Presumably they could be if the level of selection is specified in a certain fashion. Assuming we take Dawkins' line (1976/$^{2}$1989, 1982, 1995), for instance, and assert that the gene as self-replicator is the unit of selection, then it appears that a gene which programs senility or mortality may hinder the replicating function not just of the organism in which it resides (and therefore of other genes in the same body or "vehicle," as Dawkins is wont to consider the body) but also of itself as well. This need not happen if a senility- or mortality-causing gene "kicks in" after the collection of all genes in the vehicle have had opportunity to reproduce themselves. But if a "mortality gene" destroys the young organism -- the organism which has not had time to reproduce -- then it destroys itself as well and can be seen as an *internal* hindering force. The same is true of genes which do not cause immediate death but which prevent the individual organism *qua* vehicle from reproducing itself.

Necessarily, "the" unit of selection will be just one among several levels which are affected by hindering or perturbing forces. If we take the individual organism as the unit of selection and assert that environmental conditions can interfere with its ability to reproduce, the same forces can (at least sometimes) be seen to hinder the chances of other levels of life associated with that individual organism to reproduce themselves as well. When Fenimore Cooper's Uncas died as a youth, "Mohican genes" as well as Mohicans as a species were threatened. That is because Uncas's death left just one living Mohican, his father Chingachgook, who may have been already beyond reproductive age. Of course scenarios can be imagined in which the death of a single individual might help rather than hinder some level of life to reproduce itself -- a sub-individual level such as genes or a super-individual level such as clan or species, for instance. Following a shipwreck, a family of seven drifting in a lifeboat provisioned for six may be better off if tragedy strikes and one member dies sooner rather than later. Being better off in this case applies to genes, clan and perhaps even species if the six survive the ordeal which would have killed the seven.

We might go even further and claim that levels such as gene, individual, and species are reference points along an organizational continuum. These levels might, depending on the purpose at hand and on the state of sciences such as molecular biology and population genetics, be supplanted by other levels. Why not a selfish chromosome, for instance, or an egotistical molecule which uses genes as means to propagate itself? Perhaps these scenarios seem far-fetched, but the very possibility will serve to make us wary of accepting the definite article in phrases like "the unit of selection" without critical scrutiny. In fact evidence mounts that the replicative function of genes is dependent upon the normal functioning of proteins, which of course cannot reproduce themselves unless genes reproduce themselves through the reproduction of the organism (Tijan 1995). What we see, then, is perhaps better described as a "selfish regress" rather than a selfish gene. (Here it may help to remember that proteins may be said to "use" genes in order to reproduce themselves. Particular proteins are "described" by specific genes which are copied from DNA by messenger RNA. The information in the RNA is a "blueprint" for the manufacture of the respective proteins.)

To put this a bit differently, one of the ironies of Dawkins' insistence that the gene is *the* unit of selection is that his concept of the gene focuses on *replication*. His notion of a *meme* as a more general category of replicator -- embracing songs, poems, catch phrases, genes, and presumably anything which, when it can use organisms as means, will reproduce itself -- is that the replicator as object gives way to the replicator as pattern. The thing itself, i.e., any particular instantiation of pattern, is not replicated; rather its *essence* is. Songs, poems, and catch phrases are replicated in so far as their *patterns* are replicated, and the same is true of genes. Let us assume for a moment that a pattern is as much the product of what an observer attributes to an object as it is a property of the object itself. If this is true, then it is hard to imagine that there is but one level of life in which patterns recur over time. Regardless of whether time is measured in terms of generations of organisms or by some other standard, it would not be surprising if lengths of chromosomes displayed the same permutation of the same basic components, nor if individual organisms have similar characteristics, nor if groups of organisms share similarities of a kind which one could attribute only to groups. Dawkins' argument for making the unit of selection precise amounts to making our perception of pattern precise: the pattern reproduced by genes arguably remains more constant through time than the pattern reproduced at other levels. Thus one might make the argument that a certain thing which is a constituent of life at whatever level -- group, individual, or sub-individual -- is *the* unit of selection. It is also likely that this unit will be chosen to be at the sub-individual level for some purposes, else it could not be used to explain phenomena at all the higher levels. (If one posits that the group is the unit of selection, it is unlikely that one would be able to explain all individual phenomena in terms of that unit, simply because the group is composed of individuals, and not vice versa.)

When one falls into talk of the unit of selection as a pattern rather than a single thing, to say that the pattern recurs at some level is to speak not just of the objects observed, but also of the observer. In particular, when an observer insists that one pattern has priority over another in the "election campaign" for the office of unit of selection, the status of privilege has to do with perception. If we wish to explain the origin of life, we can say that an object such as a group of amino acids reproduced itself, and that the self-replicating object became ever more complex until simple organisms evolved. This is a rather startling empirical claim. When one says that patterns emerged which could reproduce themselves, that these patterns gradually became more complex but retained the ability to replicate, one might as well speak of a metaphysical principle which somehow had an effect on matter: "In the beginning there was the word (logos)...," as we read in the Gospel of John -- a focal point for the longstanding debate on how two mutally exclusive realms can have any effect on one another.

Now one consequence of the view which finds a recursive "motion" in statements about fitness and about evolutionary concepts in general is this: it is not only impossible to identify a single unit of selection, but it would be positively deadly to the theory as a whole if such a unit were ever finally, once and for all, identified. What the theory cannot accommodate is a static realm of observation. The frame, so to speak, is not large enough to encompass the entire picture and must be constantly moved about. Here is a related thesis: it is impossible to defend the claim that any entity which is defined functionally is *the* unit of selection. Dawkins himself acknowledges the impossibility of defining the boundaries of a gene by observing a physical length of chromosome *per se*. Instead we define genes by noticing the effects which changes in the pattern of chromosomes actually have. We might notice that if the pattern varies in a certain way in human beings, for instance, then eye color changes. Expressed more basically, genes as units are defined according to how their presence as patterns corresponds to various phenotypes and behaviors. Thus it is not just that certain genes are associated with specific phenotypes and with the ways in which organisms act. Much more than that, genes are *defined* by that association. Because of the definitional dependence of genes on phenotypes and behavior, we should question whether it is a truism to claim that genes are the units of selection when that statement means, in effect, that when genes are affected by the environment, they in turn affect phenotypical and behavioral phenomena.

Notice that this sort of definitional dependence does not apply to the assertion that the individual is the unit of selection. We have good evidence that individuals were defined in the case of most organisms long before the notion of natural selection was ever expressed. Individuals are, one might say, a *natural* category in the sense that we seem to perceive them independently of anything like what we would normally call a theoretical commitment (in the scientific sense). It may be that our perception of individuals is conditioned to some extent by culture, so that our ability to distill reality is learned rather than natural in the sense of being innate. (In the next chapter we will see evidence that the "folk taxonomies" of "primitive" peoples tend to track very closely with modern "scientific" taxonomies.) But I know of no example of a culture which does not distinguish individuals. The nearest thing I have found is the assertion that a culture fails to abstract upwards, that is, to perceive or to name a category which encompasses individuals or sub-categories. Along these lines Bryson claims that "[t]he aborigines of Tasmania have a word for every type of tree, but no word that just means 'tree'" (1990: 15). But let us suppose that it were not natural (that is, not culture-independent) for human beings to perceive individuals. Even so, this would be a far cry from the contention that perception of individuals is the result of a theoretical commitment. The perception of genes, by contrast, depends very much on a certain theory, and this theory is built on the relationship between genes on the one hand and phenotypes and behaviors on the other.

## 2. The unit of replication

Now the interesting thing about our discussion of the unit of selection thus far is that it applies to what we might call the unit of reproduction (or replication) as well. That is, whatever levels of life are acted upon (selected) by hindering or perturbing forces are likewise affected in their ability to reproduce themselves. Magua *qua* environmental (external) hindering force killed Uncas, presumably eroding the ability of Chingachgook, the last of the Mohicans, to perpetuate his own being as well as to reproduce his genes, his clan and even his species. The death of the seventh family member in the lifeboat furthers rather than hinders replication at the same levels. So the abstract pattern goes like this: an external force hinders or perturbs (selects) what is normally called the unit of *selection* in such a way that it cannot replicate itself as well as before. With respect to selection by external forces, the unit of selection is also a unit of replication, that is, a level of life whose ability to reproduce itself is hindered or furthered, perturbed or left unchanged, by the external force.

Thus we have a close correspondence, sometimes even an interchangeability (depending upon the context of discussion), between "the" unit of selection and "the" unit of replication or reproduction. (Given what was said above it should be clear that selection and replication act at many levels rather than at a single one, but for the moment we can tolerate the inaccuracy introduced by the definite article.) Can we trace such a close relationship between the unit of selection and what we might call the "unit of fitness" as well? One motivation for attempting to do so is probably already clear: if we can ground fitness in an empirical level of existence such as the gene or the individual organism, then we will have gone some distance toward resolving the question of whether fitness is reducible to physical properties (an issue we discussed at greater length in chapter four above). But there is another, more immediate goal which concerns us here: successfully associating fitness with whatever we take, for a specific purpose at hand, to be the unit of selection or replication will give us confidence that our talk of fitness can be precise with respect to its referent. For instance, when we assert that an individual is fit in a given environment, we will mean just that and will be correct in linking fitness to that level of life. We will not be speaking merely loosely or inaccurately or conveniently. We will not *really* mean that it is a gene which is fit and that the individual is, à la Dawkins, to be seen merely as the vehicle in which that gene rides into the future. An allegory may help emphasize the fact.

### (1) A tale of levels

Suppose at some point in the future there is a brisk trade in robots. Among the selling points -- including price, durability, how many tasks a given robot can perform, and how quickly it can be programmed, for example -- is the ability to self-replicate in common environments. "Buy our new Sophisticated Lady model," boasts an ad aired by Daimler-Benz, "and nine times out of ten you'll be the proud owner of a second Lady within eighteen months *at no extra cost*!"

The facts behind this claim are not surprising. The industrial landscape of the future era in question is rich in everything needed to produce another robot: there are chips, circuit boards, and scrap metals virtually everywhere, along with tools to manipulate them. Manufacturers program a scavenger feature into models such as the Sophisticated Lady which prompts the robot to notice the resources of its environment, collect the necessary parts, and from those parts to assemble robots like itself. The number of robots produced depends upon the richness of the environment.

(The prudent buyer will also provide work and storage areas for the Lady's convenience.)

Now we can distinguish various components of the robot. For instance, there is software which includes instructions to make the robot attempt to self-replicate. There are also many hardware components, among them ones involved in robot reproduction.

Question: What "level" of a robot's being has the drive to reproduce itself?

Answer: Any number of levels. The software which motivates replication necessarily replicates itself along with all the other aspects of the robot (but see Hampe and Morgan 1986 on viewing DNA as information under Dawkins' schema). Similarly, a circuit board critical to the reproduction task functions to re-create itself in addition to the other components of the robot. It can justly be said, as well, that an individual robot strives to reproduce itself as a whole. Further, all the copies of a given model -- all the Daimler-Benz Sophisticated Ladies, say -- might be evaluated as trying to reproduce themselves, so that there is a group inclination to reproduce, too. Or the perspective might be widened even further to warrant the nearly analytic assertion that it is all robots or even all things in general which are "programmed" to reproduce that tend to do so.

All of these statements would be true. In other words (and this is the first of the two claims promised above), one can view virtually any "level" of existence associated with reproduction as striving to reproduce itself. But no level is privileged (the second claim). That is, there is no warrant for claiming that a circuit board rather than the software or the individual robot is the unit of reproduction; no level can function independently of the others so no level actually does so.

It may be objected that a certain portion of the robot, definable as specific components and software, is indeed *uniquely* critical to the replication phenomenon. Such a *sine qua non* (the objection would continue) is indeed privileged: it, and only it, may be called the unit of replication. But given the scenario above, how would we identify such a level? Is it software? Is it the memory in which the software resides? Is it the hardware which functions together with the software? It appears that if a unique level exists in any sense, it is in so far as a certain purpose plays a role. A software engineer may see her part of robot design, construction, or maintenance as being the center of the robot's universe, so to speak, but the hardware jock down the hall may well have a different opinion.

(2) *A* unit of selection rather than *the* unit of selection

Applying the moral of the allegory to the debate over the unit of selection among organisms, it should be clear that Dawkins is right that the gene *can* be seen as the unit of selection to the extent that it can be seen as the unit of replication. This is not true of all memes. To understand this, we might speculate that in some human sub-cultures, the pick-up line is an important part of courtship. Imagine that through a process of evolution -- i.e., a process in which a fairly stable environment prompts gradual modification -- a super pick-up line is developed. With ever-increasing success this lineage of lines is used and passed on by middle-aged men as they relive past glories in the presence of their pubescent sons. Besides passing on the line, the fathers explain to their sons that the "mating game" is essentially competitive and that therefore they should avoid sharing this line with anyone who might become a competitor. Certain illegitimate children may never hear the line, but that doesn't matter; the main thing is that, at least until the environment changes (perhaps in the form of a forewarned and therefore forearmed female subpopulation), "the line" is a dominant meme. Now we can say with some justification that the line is a replicator of itself and that it tends to reproduce itself with little or no random variation. Is there a sense in which the line, like Dawkins' gene, uses human beings as vessels to reproduce it?

I submit that this is not the case, and not because the line is dependent on voluntary, intentional usage. It is wrong to consider the line to be associated with *the* unit of selection in the environment of the night club because it would be senseless to say that the line motivates men *qua* its vessels to reproduce. Whatever motivation there might be to mate and create offspring is apparently built-in by a gene or genes which can be related to some hard-to-define aspect of consciousness. We might call this aspect the reproductive instinct. For simplicity's sake, let us suppose that there is a single gene, a single length of chromosome, which causes this instinct. Then it might be appropriate to speak of "a" selfish gene which, to speak metaphorically, wants to recreate itself by motivating its host to reproduce. But the myriad other genes which constitute the host's body and cause other aspects of the host's structure and behavior by no means want the same thing (again speaking metaphorically; of course we have no reason to think that any gene "wants" to do anything in a sense univocal with the denotation of "want" in the realm of human consciousness). But *qua* replicators, don't all the genes want to reproduce themselves? Certainly, but they do not want to use the human body as a means of achieving this goal.

To make this clear, let us try to describe the case without resorting to anthropopathic language. Presumably every genome can be associated with a vast number of permutations of its alleles. What permutation actually occurs is to some extent random, but it is reasonable to believe that certain permutations are more likely to produce successful phenotypes than others. Here, success refers to survival and production of descendants. (I use "descendants" rather than "offspring" to emphasize that mere number of offspring in the succeeding generation is not a good measure of a lineage's robustness; q.v. Morris 1986.) Other, competing sets will be either more, less, or equally successful, which in turn implies that a subset of the possible permutations will, *ceteris paribus*, come to dominate the lineage (unless some other agency such as random walk comes into play).

But it is not the case (and after several readings of his major works I am not sure if this is Dawkins' position) that only the gene can fulfill the role of unit of selection. The individual and group, and other levels such as the phenotype or perhaps something "lower" than the gene apparently can fulfill the same function. In short, the gene can rightly be viewed as *a* level of selection (depending on the purpose at hand) but not *the* unit of selection. Consider the possibility that human beings will eventually be able not just to engineer their own genes and thereby their own phenotypes, but that we may regularly increase the *range* of our phenotypes. Artificial intelligence luminary Marvin Minsky (1994) speculates that parts of the human brain will one day be as amenable to replacement as hearts, livers, and some joints are today. (Interestingly, Minsky's prediction was anticipated to some extent by Teilhard de Chardin; q.v. Barrow and Tipler 1986: 199). The significance of Minsky's speculation for the debate over the unit of selection is not hard to see, but it is a bit difficult to decipher. If we augment our neural systems by computers, then arguably the creature which results -- call it *Homo hypersapiens* -- was created by selfish genes. But it is not hard to imagine a scenario (fit to be the plot of a science fiction novel) in which succeeding generations of *hypersapiens* supplant increasingly more of their neural systems and other phenotypical attributes with non-organic parts. Those living tissues which remain will be engineered so that the metaphorical struggle of gene against gene, or allele against allele, is effectively ended. Genes, in short, will be disenfranchised. Purists in the selfish gene camp will attempt to argue that *hypersapiens*, even if it were to become virtually inorganic and its individual members practically immortal, is ultimately the product of genes which competed for hegemony in the distant past.

There is no need to argue the point. The human genome as of 1997 (or any other arbitrarily chosen point in time) almost certainly did not always exist. If our hypothetical descendants, *hypersapiens*, can be said to have evolved from the genome, then the genome must also have an ancestor. It evolved from simpler organisms with fewer genes, and in some sense both levels -- genetic and organismic -- as well as in between levels (e.g., phenotypes) competed against one another as well. How did these simpler genomes and phenotypes arise? Eventually we will run out of simplicity, so to speak, so that genes can no longer serve as the explanation of evolution's spectrum.

A more profitable approach is simply to treat no "level" of life (e.g., genetic, phenotypic, organismic) as privileged, as *the* unit of selection in an absolute sense. Rather, we can choose to equate the unit of selection with the unit of fitness -- as the thing which happens to be most closely tied up with differential reproduction for a given researcher's particular purposes. We can then treat fitness, in turn, as a recursive function.

## 3. The Substrate of Evolutionary Change

### (1) What Changes?

The question can of course be taken in two ways: (1) What thing changes? (2) For a given thing, which of the changes it undergoes are interesting to us? Let us consider the first of these questions. The basic problem is to decide which of the myriad elements of an organism's structure and behavior are of interest. This amounts to inquiring after the foundation of a taxonomy. We could call this the Moby Dick dilemma, and not just because the issue constitutes a whale of a problem but also because the fictional narrator of the novel takes on Linnaeus on the issue of how whales are to be categorized. Acknowledging that the creatures are warm-blooded and have lungs, the tale's narrator nonetheless refuses to deem whales to be mammals and instead maintains his own definition: "[H]ow shall we define the whale, by his obvious externals, so as conspicuously to label him for all time to come? To be short, then, a whale is a *spouting fish with a horizontal tail.* There you have him" (Melville 1851: Ch. 32: 140). Taking Ishmael as a researcher of a kind, or at least an interested layman, we can see this remark as exemplifying the tension between different species concepts.

Gould makes the argument that species are real on the basis of anthropological evidence. This argument is based on the proposition that non-Western cultures recognize more or less the same delineations among animals with which they come in contact as the Linnaean system would see. One must allow for some variation based on the interests of the group doing the classifying (e.g., a tribe), but the similarities with scientific classification are striking. A representative observation comes from a leading expert on speciation.

> Ernst Mayr himself describes his experience: 'Forty years ago, I lived all alone with a tribe of Papuans in the mountains of New Guinea. These superb woodsmen had 136 names for the 137 species of birds I distinguished (confusing only two nondescript species of warblers).' (Gould 1980: 207)[46]

The comparative-cultural argument for the reality of species, then, goes like this: Westerners influenced by Linnaeus and the ensuing scientific tradition recognize essentially the same species as peoples who observe animals closely but who have never heard of Linnaeus and who have escaped the influence of Western science. This shared, cross-cultural delineation of species could only occur if species had a reality apart from the biases of the observer. "Unless the tendency to divide organisms into Linnaean species reflects a neurological style wired into all of us (an interesting proposition, but one that I doubt)," says Gould, "the world of nature is, in some fundamental sense, really divided into reasonably discrete packages of creatures as a result of evolution."[47] Gould does not examine the notion that species may be described by most classification systems according to the purposes of the observer. If there were enough commonality among the respective purposes of different cultures, then the similarity among delineations would be predictable rather than surprising. There may in fact be such a common purpose: Western scientists want to catalog regularly recurring features so that they have enough stability to get on with whatever interests them. Similarly, the hunters of a remote tribe want to identify such regular features so that they can begin to generalize about where to find what game and when, what tastes good, what fights back, what's poisonous, etc. This is reminiscent of Pinker's explanation of species concepts as those ways of distilling nature which are profitable (pp. 341 - 342 above).

For the sake of clarity, perhaps it would also be well to say what most models of evolutionary change do not represent (at least not necessarily). Progress is arguably a central, necessary notion in the 18th century roots of evolution, and even now progress is integral to the thought of key evolutionary thinkers such as E.O. Wilson in the field of social biology (Ruse 1995: 122-137). But so-called progressionism is denied and even attacked by other thinkers. How do we represent progress to ourselves? It seems clear that *increase* of a kind is representable by tree-type illustrations. Not increase of some difficult- or impossible-to-define, abstract quality such as complexity, but rather a simpler, more basic kind of increase. Perhaps we could call this an increase in *diversity*, though presumably the simple number of living things grows as well. The typical picture starts with one or a small number of vertices, depending on the illustrative task to be fulfilled, and then the "tree" grows: the lines lengthen and the number of "branches," of vertices, steadily increases as one proceeds in the direction which represents time. It is interesting that the increase is not just perceivable but countable in this representation (the lines are so many units long, and there are a specific number of vertices in the tree as a whole).

That discrete nature, the countability, is a very interesting feature, one which goes far beyond the concepts of change and increase in diversity which are obviously central to evolution. A simple increase could be represented by a single line of positive slope imposed on an ordinate representing diversity and an abscissa showing time (or vice versa). Such a picture would show that life has become more varied, or in other words that life's spectrum has become broader, without implying that there are well-defined divisions among points on the spectrum at a given time. By contrast, the discrete nature of the branching tree corresponds to the notion that the variety of life is divisible according to two aspects. The first of these is lineage, which relates a given organism at a given time to other organisms in the past (the organism's ancestry) and in the future (its progeny). Consistent with the terminology of linguistics, we might term this relationship among organisms *diachronic* (q.v. Lyons 1981: 45 ff.). The one-way progress of a line in either direction represents this relationship between ancestors and descendants. The second sort of divisibility implied by the branching-tree picture is again a relationship among organisms, but this time it is *synchronic* -- based on comparing organisms which exist at the same time without regard to lineage. Two synchronic criteria of comparison are structure and behavior. The first, structure, is at once natural and problematic. It seems natural to place organisms which share certain features -- shape, size, coloration, for instance -- in the same conceptual category. However, it is not clear that there is any privileged way of selecting which features of organisms to treat as defining aspects (Is a difference in coloration sufficient to demarcate species?) and what standard to use in judging the selected aspects (How different must coloration be to demarcate species?).

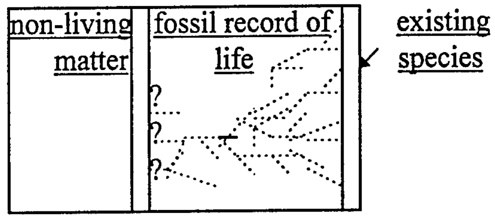
If there is any privileged way of conceptually dividing life, then that method is better than other ways only with respect to the observer. By this is meant that a given observer may have cause to focus on certain aspects of organisms and ignore others. This is true for two reasons. The obvious one is that an observer can be purposive. He can have reasons for singling out particular features of organisms, and these features can in turn form the basis of a taxonomy. A naturally deaf organism would not care how the organisms around it sound, but it might have a great interest in their relative sizes. Besides determining which organisms are a threat, size has a great deal to do with what's on the menu. So here we have the possibility of a taxonomy which is privileged in a relative sense. A related possibility arises if we consider what Aristotle might have called the *arete* (virtue) of a given organism. We humans make a germane example in this regard, since after all we are focusing on the tree diagrams which are so frequent in our theories of evolution. What is out particular strength? Or phrased more neutrally: What aspects of ourselves seem unique, either absolutely or as a matter of degree? Reason, is the classic answer, but others -- based on ability to use sophisticated language or make intricate tools, for instance -- make sense for other purposes. But for the naturalist, before the world is the plaything of reason, before it is the subject of conversation, before it is an object to be manipulated with tools, it is a thing to be observed, that is, sensed. We might well expect, then, that a naturalist's taxonomy of living things would be based on easily collected sense data first and foremost. That leaves a lot of possibilities open, but the range is narrowed if we assume that a taxonomy based in perception will derive from the observer's keenest senses, less from the others. (What constitutes a "keen" sense is difficult to define without appeal to reason and language. It seems that we humans are much more adept at reasoning about things we see, hear, and touch, for instance, than about those which we smell. Not that olfactory stimuli aren't powerful and ubiquitous in our lives; on the contrary, they move us with a power not surprising given our evolutionary history. But if a taxonomy is developed *from* perception, it is nonetheless constructed *for* understanding. We want to understand more about our world with the aid of whatever taxonomic principles we choose. Therefore we base our taxonomy primarily on what is arguably our keenest sense, namely, vision. It is shape, coloration, and visible behavior which guide us. Thus the two senses in which a taxonomy may be privileged -- according to the purpose and perceptual forte of an agent -- are not absolute truths about the world.

It may be that a third criterion exists in the notion of function. If organisms which appear to be very different in structure and behavior nonetheless have parts which function similarly, the similarity of function may be used to group the organisms together at some level and for some purpose. One might conceive of function in this sense as being a special case of behavior. "Behavior" in its normal sense refers to how autonomous units such as organisms act, while "function" in the sense used above describes how parts of organisms act. This is worth mentioning because, as we will see, it is not clear what constitutes an "autonomous unit." The traditional view is that organisms as well as groups of organisms can be described in terms which attribute purposiveness. More recently, it has been suggested that parts of organisms can (or perhaps should) be viewed not just as functioning things but also as behaving things as well. Dawkins' metaphorical language -- depicting the gene as "selfish" and organisms as "survival machines" manipulated by genes -- is perhaps the most emphatic along these lines.

The diachronic and synchronic aspects blend in reproductive behavior. Some organisms are very nearly identical structurally, but because they do not form part of a breeding population, they are not considered to belong to the same species. This can be true even though two individuals are capable of breeding but, in nature, do not (cf. the account of *Chrysopa downesi* and *carnea* in Ridley 1983). This fact is in accordance with the best-known species concept, generally attributed to Ernst Mayr.

This definition of species holds that the members of an interbreeding population constitute a species; individuals outside this independent group are excluded from the species relationship. In its most extreme form the *cladistic* philosophy of taxonomic grouping relies wholly on interbreeding: two organisms are related only to the extent that they share common ancestry.

A basic model of evolution can be offered in the form of a simple diagram.



One can find pictures similar to this one in numerous expositions of evolutionary biology (e.g., Ridley 1985: 3, who offers three graphic models of evolution assuming, respectively, a single ancestor, multiple ancestors, and instances of special creation; Futuyma 1983 and 1986; Simpson 1949; and of course, Darwin 1859, to name just a few). Obviously all such pictures represent change, which is a concept unifying the various theories of evolution. If one accepts that science, presumably including evolutionary biology, proceeds by describing the interaction of entities and forces, then our simple picture should prompt us to ask three major questions about change as the central theme of evolution: (1) What is it that changes? (2) What results from the change? (3) What causes the change? (We won't attempt to answer the third question in this chapter.)

In other words, we want to understand the diagrammatic endpoints, so to speak, as well as what connects them. If we take the whole picture or any segment of it large enough to represent more than one organism, then we can observe an initial status represented by one or more organisms as well as a terminal status represented by an individual or several organisms, and the two are connected by a line or lines which chronicle the effects of some force or forces.

### (a) The first question: What is it that changes?

This first of three questions about change tends to be upstaged by a different but related question: What is the unit of *selection*? The goal of posing this common query is to discover the part or unit of life which is the object affected by the force of natural selection. Certainly this is a critical issue within the theory of evolution, but there is no universal consensus on how to answer the question. We, too, have briefly considered the matter and how it is affected by a recursive understanding of fitness. But in the spirit of trying to *describe* a phenomenon before attempting to *explain* its cause, perhaps we would do better to inquire into the "unit of *change*" before we attempt to specify the unit of *selection.* Hence the first of the three questions about change on our present agenda. As it turns out, the investigation into what changes to a large extent parallels the analysis of what is selected (i.e., what is the object of natural selection). But there is a difference between the two questions if we assume that something more changes than is actually selected or not selected. In fact this appears to be the case. All of the entities which are candidates for natural selection seem to change: genes change in fact (in their material being) if not in pattern; individual organisms change; so do groups of individuals. Unless all of these levels are taken as "the" unit of selection (something which no one has proposed, so far as I know), that which changes cannot be equated with what is selected.

What we are after is the stuff which characterizes and makes sense of the basic notion of evolution -- change -- and we can start by thinking in Aristotelian terms about the "material cause" of change. Our present purpose does not require a lengthy exposition of the distinction among material, efficient, final, and formal causes (see pp. 39 - 49 above for a fuller treatment). Suffice it to say that a prerequisite of tangible change is some material which is transformed (the material cause) by some force or by an object which serves as the mediator of a force (efficient cause). Although most evolutionary biologists would reject any analog of formal cause (essential pattern) or final cause (goal or target state) in a long-range sense -- that is, in the sense of a perfection principle -- it seems clear that to the extent that reproduction at any level (gene, phenotype, individual, or group) is replication of a previously existing pattern of physical qualities, that pattern might be seen as both the formal and final cause of reproduction.

For evolution to take place at all, something must be transformed into something new. But what? If we take any two points on a lineage, the organisms represented by the points bear a temporal and causal relationship to one another. But suppose we ignore these aspects. In that case, any interesting connection between two organisms belonging to a lineage is a relationship of similarity. Now likeness can occur (or fail to occur) on more than one level. For instance, organisms may share common genes without also having all phenotypes in common. But suppose we concentrate on a single level only. It does not matter which one. What we need is a criterion of inclusion, a reason for claiming that two organisms are so much alike in some respect, that they should be grouped together for some purposes. Three questions are central here: Which aspects of organisms are important in questions of inclusion? How much alike must two organisms be with respect to some quality in order to justify their inclusion in the same category? What purposes are relevant in questions of inclusion?

For each of these questions there is a range of possible answers. Although it would be interesting to consider each of these answers in some depth, it is not necessary at this point. It suffices to note that the first question can be correctly answered by appeal to any level -- genetic, phenotypic, individual, group -- depending upon how the third question is answered. Similarly, the "right" answer to the second question depends upon how we answer the third question. To summarize in slightly different terms, the aspects which we choose to deem important in processes of change, and the degree of similarity which we demand among chosen aspects, depend upon our purposes. Dawkins rightly points out that genetic relationship may alter our perception of apparent altruism at the level of the organism. But from such examples it would be wrong to conclude that the gene is *the* unit of change -- the part of the spectrum of change which is most important. It seems most profitable to analyze other phenomena by concentrating on individual organisms rather than genes.

At this point it might be well to ask just what is meant by a replicator (a term of which Dawkins is extremely fond). Dawkins implies that a replicator is something which, metaphorically speaking, *wills* replication of itself. By this is meant that the gene is programmed to reproduce itself *ad infinitum*; unless it is stopped by some limiting condition, that is what it will do. But normally it cannot do this alone. In the case of organisms which reproduce sexually, a given gene can reproduce itself only when its host organism cooperates with another organism to begin the process of meiosis. Suppose we argue that if A is the only means to B, then to attempt to do B is to attempt to do A. If this is true, then to say that all genes try to reproduce themselves is to say that all genes try to cause their host organisms to take whatever steps are necessary to start the process of meiosis. But such a claim would necessitate that whatever other effect every single gene might serve to bring about, alone or in combination with other genes, it must also motivate its host organism to mate in some fashion. It is not clear to me how this claim could be decisively tested. Perhaps it would be possible to localize the "mating instinct" in a number of genes which, if somehow removed or neutralized, would render an organism incapable of mating or unmotivated to do so. If the host organism had no inclination to mate despite the fact that the vast majority of its genes are normal, presumably these genes could be considered inert or neutral with respect to the mating instinct. This status would in turn make the label "selfish gene" inappropriate. As was shown above, the implication of the phrase is not just that every gene is a replicator of itself under the right conditions, but that *every* gene is somehow programmed to bring about those conditions. To repeat, this last claim seems highly dubious.

By Dawkins' way of reckoning virtually all conditions which limit the replication of the gene are environmental. Even the body in which the gene resides becomes a mere means for the propagation of selfish genes. But this view of the gene is based on the assumption that replication is somehow what a gene automatically does in the absence of any obstacle. For the reasons suggested above, this view is false. The average gene is in no sense, metaphorical or otherwise, inclined to replicate. It would be better to call the gene a *means* of replication. In this sense a gene might still be called a replicator, but the same term could apply to the motivator of replication as well (wherever we might choose to ground that function). This seems to conform to common linguistic usage, where the agent who wills to do something and the tool which the agent uses to carry out his wish can share the same name: a shooter and his six-shooter, or a couple of hikers (in the sense of backpackers) and their hikers (their boots), for instance.

Such a perspective should lead us to consider whether the motivating force for replication is not essentially extra-genetic. We can approach this issue by asking why Dawkins denies the status of replicator to other levels of life. Consider an individual organism which reproduces sexually. Presumably it will be impossible for such an individual to reproduce itself exactly. At best it can hope to reproduce some of its phenotypes in some of its offspring. Thus at first glance there seems to be a strong difference between gene and individual organism in this regard. Yet the strongest proponents of the selfish gene admit that the gene is a somewhat slippery concept. Physically speaking a gene is a length of chromosome which may or may not succeed in reproducing itself in any succeeding generation, since a "copying error," mutation, is always possible. And as already argued above, the gene cannot do this alone nor does the average gene tend to do so.

In fact neither the individual nor any level of life below that of the individual organism can justifiably be called a replicator in the sense of self-replicator. What is replicated at these levels? The individual cannot replicate itself; the average gene does not of itself tend to. If there is a level at which anything is replicated, it is at the level of a group of organisms which share some set of qualities which (again depending on the observer's purpose and predilections) serve to group them together and whose members perpetuate those qualities through reproduction. (Again we are speaking of sexual reproduction here; the case of asexual reproduction will be treated later.)

By this point the reader will have noticed that our discussion seems to be drifting more in the direction of shared qualities than of aspects which do *not* remain the same across generations of sexually reproducing organisms. In other words, the emphasis seems to be more on continuity than on change. This is no accident, and in fact, although the two -- continuity and change -- seem to be opposed to one another in common speech, in our present context they are closely allied. In observing the changes across generations of organisms we need some sort of standard of comparison. We are, after all, concerned not just with *any* changes when we speak of the evolution of a lineage of organisms, but rather with changes in those particular categories which unite individuals in a common lineage. It is these categories which provide a continuity overarching the changes. A particular quality can be manifested with some variation yet remain within acceptable boundaries, as mice can be big by mouse standards or small by mouse standards but not as big as beavers. In short, change (or at least *recognizable* change) must coexist with enough continuity to justify uniting separate individuals under synchronic umbrella concepts such as species and diachronic categories such as lineages. When we seek the unit of change, then, we are apparently on the trail of something changeable which coexists with constant elements.

Whatever transformation takes place does so across generations (if we accept a Darwinian rather than Lamarckian version of the theory). But if the change is across generations, then one thing which has changed is the individual organism. Whether a kind of organism reproduces sexually or asexually, there is at least a temporal distinction between the parent and offspring organisms. Along with change in the organism necessarily comes a change in the constituent parts of the organism as well. Although it is convenient to speak of an organism receiving the genes of its parents, this is of course only a figure of speech. (Here the plural form, "parents," implies that we are speaking of sexual reproduction, but the same points apply to asexual organisms as well, *mutatis mutandis*.) The form of the offspring's genes may be very like that of the parents', but the genes are different. Even among the half or so of the genes which were had from one parent, what is shared is pattern rather than the material in which the pattern is made manifest.

Now similarity of pattern or form can be observed not just in the inheritance of genes but also in the passing-on of the qualities which the genes cause. Phenotypes are essentially patterns rather than things in the same sense that this is true of genes. When one speaks of two organisms sharing some phenotype -- a long neck, say -- one does not mean that the two organisms are joined behind the ears and must go through life with just one neck between them. Rather, there are two *similar* necks. This may seem a trivial point, but perhaps the convenience of speaking about shared traits as though "trait" referred to an object rather than a pattern has caused pointless debate over issues such as what "the" unit of selection is. Dawkins, for instance, works his theory hard to find enough continuity across long periods of time in order to make sense of certain observed phenomena. For instance, it is difficult to reconcile a *laissez faire* morality with apparent instances of altruism among individuals. However, if we assume that the source of activity -- the control center, so to speak -- is a pattern which seeks to increase itself, then selfishness can be attributed to the pattern itself while (illusory) altruism belongs to the realm of organisms conceived as mere means for the perpetuation of patterns. If one can understand phenotypes as being patterns every bit as much as genes are patterns, then it might seem possible to claim that phenotypes are in fact the units of selection, the entities which "intend" (metaphorically speaking) to replicate themselves. But thinkers such as Dawkins give priority to genes apparently because this level of pattern is always evident, in the fetus as well as in the adult (Dawkins 1976/$^2$1989, 1982).

Perhaps continuity is a reasonable criterion for preferring the gene to other units of selection, but is there another criterion which would motivate us to accept another "level" of life as the unit of selection? In other words, is there some central aspect of evolution as change other than continuity? Those who oppose accepting the gene as the unit of selection quickly point out that natural selection as a force acts most immediately on phenotypes rather than on genes (Dennett 1995). It is inescapable that because of the causal connection between the gene and the phenotype, natural selection must act on both if it acts on one. But the salient point is that the gene itself does not interact as *directly* with the environment as its phenotype does. We will need to refer to this quality of the phenotype again, but the literature seems not to contain a specific name. We will have to invent one. If the gene has a greater continuity than the phenotype, let us say that the phenotype has a greater *proximity to force* than the gene. By this is meant that the force which animates Darwinian evolution, natural selection, acts immediately on phenotypes, which in turn affect genes. Seen from another perspective, a particular phenotype stands between the allele (or more likely a *combination* of alleles which caused it) and the force of natural selection. (Of course no perfectly sharp dividing line can be drawn between the gene and phenotype in this respect, particularly in cases where acquired characteristics are in fact seen to be heritable, consonant with Lamarckism (Hill 1967). We will study this point in greater depth below. In order to understand the basic positions in the debate over the unit of selection, however, let us accept for the moment that natural selection "edits" living things based on characteristics like long legs versus short legs or keen eyes versus weak eyes, regardless of the genetic basis of such traits.)

What characterizes the debate over the unit of selection, then, is an argument based on which aspect of change -- continuity or proximity to force -- is seen to be more important within the overall theory of evolution. Dawkins and other advocates of the selfish gene see continuity of pattern as having precedence; Gould and the camp of pro-phenotype thinkers attribute greater importance to the efficacy of natural selection, which in turn means that they emphasize the unit which seems to be the direct object of natural selection. (It is ironic that Gould is criticized so severely by some -- Ruse 1995 and Dennett 1995 are recent examples -- for being a revisionist in rejecting gradual evolution in favor of punctuated equilibrium. Yet Gould remains a traditional Darwinist in as much as he emphasizes phenotypes over genes and focuses on natural selection more than on continuity.)

At this point it might be suggested that the unit of selection must be some Aristotelian *form* which remains constant among various individuals. When we say that children inherit the genes of their parents, for instance, we do not mean that some object called a gene passed on to the descendants. We mean rather the gene as replicator, capable of constructing not just the same phenotypes as the parents but indeed the same genes in the sense of genetic patterns -- ordered subsets of amino acids within an ordered set called a chromosome. The form is clearly key. Or perhaps the form which remains constant is a concurrence of phenotypes, or macro-characteristics.

Before analyzing this view more closely, it would be well to refine our understanding of how phenotypes differ from genotypes. Genotypes might be seen as qualities of an organism just as phenotypes are. The connotation of phenotype leads one in the direction of a "macro-quality" as opposed to the kind of quality which a gene represents. The modifier "macro" is intended to help distinguish between qualities such as a tiger's stripes or a giraffe's long neck and the genes which cause those qualities. When one pauses to consider, it is not at all clear what the qualitative difference between geno- and phenotype *as effect* actually is. A difference is only clear in so far as the genotype is seen as the ontogenetic and phylogenetic cause of a certain phenotype. In other words, within a given organism, a certain gene or combination of genes "causes" a phenotype. But the gene or combination of genes is itself caused, phylogenetically, by earlier-existing genes or by phenotypes or individuals, depending on our perspective.

Perhaps it is individual organisms that change, but if the individuals constituting a type change, then the type itself, the group, changes. And by the same reasoning, there may be a substrate of component parts below the level of the integrated organism which likewise changes. Or the organism may be divided conceptually, so that change affects *conceptual* parts rather than "real" ones. To repeat, an easy way out of the problem of selecting *the* unit of change may be simply to appeal to purpose: although a spectrum of units can in fact be said to change, the level which is taken to be the axis of change is specified according to the goal of research. A paleontologist may interest himself primarily in changes among groups of various sizes (e.g., phyla, classes, families, genera, species), while a molecular biologist may need to focus on the gene or an even smaller part of a chromosome. Purpose is the key. At each level there are categories of continuity, that is, qualities which must lie within certain tolerances. Within these boundaries variation is possible. From one perspective, these variations are the units of change which we have sought. So long as the boundaries of the categories are respected, we can say that change occurs within a taxon. There is no need to posit the emergence of a new kind of organism so long as the categories remain intact.

So much for the kind of unit of change which amounts to a unit of continuity. What about the unit of change in a purer sense? Clearly there are many aspects of organisms which mature -- entire individuals, phenotypes, and other conceptual segments of life such as the elasticity of skin. Such ontogenetic changes make poor units of change. Naturally, allopatric speciation is in some sense a function of variable rates of maturation of features within single organisms, but the major changes which characterize such speciation are best analyzed in terms of categories of continuity. If we choose an example of allopatric speciation, we will find that when the quicker or more complete maturation of one aspect of an organism with respect to another aspect results in speciation, it is because the boundaries of continuity are broken. In other words, we can observe a number of changes taking place in organisms during the course of their evolution, but all the interesting changes can be characterized in terms of the categories of continuity.

One final note before turning to the next basic question suggested by our pictorial model of evolution above. Evolutionists frequently use the development of human languages as an analog to the biological evolution of structure and non-linguistic behavior (Futuyma 1983, Gould 1970). There is clear evidence that pre-Darwinian thinkers were indeed able to deduce much about the history of human language based on the assumption that peoples and cultures which Europeans held to be basically inferior to themselves and their own culture spoke languages which somehow evolved into our own. One example is Sir Alan Jones' analysis of Sanskrit (Bryson 1990). Along the same lines Ambrose notes that

> [President Thomas] Jefferson had a passion for Indian language, believing he would be able to trace the Indians' origins by discovering the basis of their language. So the gathering of vocabularies was an important charge on [Captains Meriwether Lewis and William Clark, who were about to set off in search of an all-water route across North America to the Pacific Ocean]. (1996: p. 203)

Lewis gathered a number of vocabularies as instructed. Indeed, he was able to deduce a common origin from rough similarities. (He had no training as a linguist and spoke English only; he relied on interpreters even for French; ibid.: 298.[48])

### (b) The second question: What results from change?

The second of our basic questions -- What results from change? -- can be taken as articulating a pivotal issue in taxonomy. The very title of Darwin's work, *On the Origin of the Species*, tells us that perhaps *the* central concern of evolutionists must be the phenomenon of speciation. Speciation in this sense is a euphemism for a concept we could also encapsulate in a phrase: Breaking the boundaries of the categories of continuity in such a way that change "overpowers" continuity. In order to make any headway in this area of research, one must presumably know what it is that constitutes a species. This in turn boils down to knowing two things: the categories of continuity (what characteristics define a species) and the boundaries of each category (what degree of variation within a certain category is possible). For instance, we may define a given species of insect, a lacewing, by its color and the time at which it breeds. Certain lacewings are dark green, others are light green. In nature, no "blended" coloration is possible, while laboratory conditions have produced intermediate tones. How a given taxonomist would handle such phenotypes depends on the boundaries of the category coloration. In this case coloration deserves the title of category of continuity because it is one criterion used by taxonomists to delineate species. But the boundaries of the category could accommodate only two basic possibilities. That was sufficient in nature but insufficient in the laboratory environment. (Recall the case of *Chrysopa carnea* and *C. downesi* described in Ridley 1985). The moral of the story is that the result of change is theory-laden in its systematization: to organize our observations, we must establish the boundaries of our categories of continuity. Such boundaries are not necessarily "given" by observation, as the example of the lacewings demonstrates.

We should also note a frequently occurring source of tension in the work of those who analyze the philosophical foundations of evolutionary biology. Evolution involves change, and not just the cyclical change of individual organisms as they mature and die (ontogenetic change) but also the linear change of entire species of organisms (phylogenetic change). In other words, among generations of organisms we see the same stages repeated over and over, so that individuals change even when a species remains stable so far as we can tell. That is a cyclical, ontogenetically-based kind of change. But of course species change as well. Such whole-species change -- which we can equate with the concept of evolution for our present purposes -- *might* be cyclical, that is, it *might* tend to oscillate through the same structural and behavioral patterns over huge periods of time. Whether or not this is true and whether or not human faculties could perceive it, we can say with fair certainty that the evolution of species *looks* linear. It seems as though species progress through phases to which they never return, particularly in cases of extinction. Because such apparently linear change is at the heart of what we mean by evolution, it has become fashionable to characterize Darwinian evolution as the antithesis of the concept of Platonic forms and to insist that Darwin's insight depended on his rejection of the forms. Mayr's sentiment in the foreword to a facsimile edition of the *Origin* is typical of this view: "Darwin started from a new basis by completely eliminating the last remnants of Platonism, by refusing to admit the eidos (Idea; type, essence) in any guise whatsoever" (Mayr 1964: xi; similar statements can be found in Futuyma 1982, 1986).

Perhaps Mayr and Futuyma overstate. How could we recognize evolution as linear change within species, even to the extent of speciation, unless we first had at least a rough and ready means of defining species? Any such means must recognize certain recurring similarities among organisms, similarities so striking and so regular that they define "natural kinds." Those similarities, in turn, can be identified with a constellation of concepts which Mayr, Futuyma and others associate (rightly, I think) with Plato's notion of *eidos*. To recognize species as eternal in a way which some associate with classical perspectives, particularly Aristotle's, is going too far. But the other extreme is just as dangerous, maybe even more so. Commitment to the eternality of the species would fly in the face of data best explained by positing the evolution of species, including speciation events and extinctions. But at least in so far as such a commitment includes the notion of *eidos* as an organizing concept, a means of identifying kinds, taxonomy is possible. On the other hand, if we reject *all* regularity, we have no chance of organizing organisms into kinds so that the notion of linear change in the structure of those kinds -- the notion of evolution, in short -- is possible. To put the matter succinctly, perceiving linear change among species requires us to recognize cyclical stability among generations. Paradoxically or not, evolutionary change is meaningless unless there is a background of stability.[49]

### (2) Mayr's distinction between typological versus population thinking

In 1975 Mayr published an essay which he believed to be, in its original formulation as a paper delivered in 1959, "the first presentation of the contrast between essentialist and population thinking, the first full articulation of this revolutionary change in the philosophy of biology" (in Sober 1984c: 14). Mayr asserts that before Darwin, "typological" thinking predominated. This view of organisms, says Mayr, was based on Plato's notion of *eide* as the unchanging realities behind the multiplicity of varying individuals which necessarily fall short of their associated perfect forms. Among the consequences of this way of thinking, according to Mayr:

> Since there is no gradation between types, gradual evolution is basically a logical impossibility for the typologist. Evolution, if it occurs at all has to proceed in steps or jumps. (Sober 1984c: 15)
>
> The observed variability [among individual organisms] has no more reality than the shadows of an object on a cave wall, as it is stated in Plato's allegory. (ibid.)
>
> The typologist stresses that every representative of a race has the typical characteristics of that race and differs from all representatives of all other races by the characteristics 'typical' for the given race. All racist theories are built on this foundation. (ibid: 16)
>
> For the typologist everything in nature is either 'good' or 'bad,' 'useful' or 'detrimental.' Natural selection is an all-or-none phenomenon. (ibid.: 16 - 17)

This all-or-nothing character, says Mayr, causes the typologist to reject evolution by natural selection.

Mayr's knight in shining armor is, of course, the populationist thinker, whose assumptions are said to be "diametrically opposed" to the corresponding perspective of the typologist. But despite what Mayr claims for this perspective, it seems clear that all of the allegedly populationist positions could equally well be attributed to one committed to a typology based on a Platonic theory of forms. For instance, Mayr claims that by the populationist's way of reckoning, "[a]ll organisms and organic phenomena are composed of unique features and can be described collectively only in statistical terms" (ibid.: 15). Mayr further asserts that the statistical characterization is a mere abstraction, so that "only the individuals of which the populations are composed have reality" (ibid.: 16). To claim (as Mayr does) that the typologist takes only forms to be real, while all else is illusion, ignores the obviously critical distinction between the ontological and epistemological positions adopted by Plato and those who endorse his theory of forms in one fashion or another: the forms, however real, cannot be completely known, so that individuals and types are indeed real objects given our epistemological limitations. But hermeneutics is not our business here. Suffice it to say that a scheme of classification based on ideal types does by no means deny the reality of individual variation. The reckoning of the (perhaps unstated) ideal which defines the type is based on observation of individuals, just as in the case of what Mayr calls populationist thinking.

Leaving aside the question of whether Mayr overstates the distinction between populationists and typologists, it is interesting to question what kind of species concept a full-blown populationist perspective would allow. It seems immediately clear that if the populationist can group individuals only by "unreal" statistical abstractions, then the resulting categories are equally abstract and unreal. Although Mayr does not examine the issue in depth in this essay, it appears that under the populationist schema, species and other taxa are established functionally. In other words, depending upon the task at hand, a researcher may decide to consider a certain collection of individuals to be of a given type or not. However, a new direction of research might well necessitate the dissolution of the current taxa in favor of another grouping.

Mayr is well known as a defender of a view of species as geographically separated, interbreeding groups of individuals (1940). How does one square this with his populationist views? Here we should be careful in order to ensure that this common "picture of evolution" does not lead us astray. In the spirit of caution, let us accept that we can view any scientific theory as consisting of two basic components, forces and the entities upon which the forces act. One of the forces which (most agree) plays a role in evolution is natural selection. Now what are the things which are affected by natural selection? This is a delicate question, one which in a slightly different form is key to the unit-of-selection problem. Briefly, this problem seeks to specify the object -- perhaps the gene, or maybe the individual organism, or yet again perhaps a group of organisms such as a breeding population -- upon which natural selection acts. As we have already discussed, each of these choices has an amorphous aspect which makes it difficult to assert that it is the object which is selected. The best-known champion of the gene as the unit of selection, Dawkins, admits at the outset of his major work on the subject that it is difficult to define exactly what a gene is. He invents the concept of "genetic unit," which "is just a length of chromosome, not physically differentiated from the rest of the chromosome in any way." Moreover, this genetic unit "will overlap with other genetic units. It will include smaller units, and it will form part of larger units" (1976/$^2$1989: 29).

My point in quoting this passage is not to question the rigor of a theory whose primary agent is not physically recognizable. Certainly it is not necessary to spear a bit of matter on a probe or isolate it under a microscope to do good science. But it is important to recognize that the concept of a selfish gene is essentially typological rather than populationist in Mayr's terms. This should again lead us to question whether Mayr overdraws the distinction.

### (3) Continuity/Stasis

Maynard Smith and Price's concept of an evolutionarily stable strategy (ESS) offers a formal account of why there is any stability at all among the behavioral patterns of organisms (Maynard Smith and Price 1973). The concept of an ESS

> can be crudely encapsulated as a strategy that is successful when competing with copies of itself'. (Dawkins 1982: 120)
>
> If a program or strategy is successful, this means that copies of it will tend to become more numerous in the population of programs and will ultimately become almost universal. It will therefore come to be surrounded by copies of itself. If it is to remain universal, therefore, it must be successful when competing against copies of itself...." (ibid.)

The immediate goal of "competing" is maximal reproductive success in Dawkins' understanding, but it should be noted that an ESS can be understood as a behavioral pattern which serves the goal of maximal reproduction only indirectly. Thus Gibbard:

> Consider now human beings evolving in hunting-gathering societies. We could expect them to face an abundance of human bargaining situations, involving mutual aid, personal property, mates, territory, use of housing, and the like. Human bargaining situations tend to be evolutionary bargaining situations. Human goals tend toward biological fitness, toward reproduction. The point is not, of course, that a person's sole goal is to maximize his reproduction; few if any people have that as a goal at all. Rather, the point concerns propensities to develop goals. Those propensities that conferred the greatest fitness were selected; hence in a hunting-gathering society, people tended to want the various things it was fitness-enhancing for them to want....Propensities well coordinated with the propensities of others would have been fitness-enhancing, and so we may view a vast array of human propensities as coordinating devices. Our emotional propensities, I suggest, are largely the results of these selection pressures, and so are our normative capacities. (Gibbard 1990: 67)

It is a commonplace to assert that evolutionary biology (at least in the modern synthesis) explains the origins of life and the evolution of species in wholly mechanistic terms, without the ghost of a reference to teleology. (Dennett 1995 is a forceful case in point.) By such a mechanistic account, the current status quo of life on the planet exists not because non-living matter and, later, organisms were in some sense *destined* to reach this point on their way to God knows what final state of perfection (assuming existing organisms have not already reached that final state), but rather because a combination of mechanistic forces and chance have acted on successive generations of life over millions of years. Thus it is claimed that the journey had no set destination, but took its direction based on the circumstances of the moment.

The question immediately arises, Is it consistent to say that a progression is determined to any extent by a factor operative at the outset -- environment, say -- and at the same time to claim that the destination of the progression is not at least semi-determinate? Consider an initial state $S_1$ in some process. $S_1$ is acted upon by forces $F_1...F_n$. If there is *any* regularity in the interaction of the forces with $S_1$ and successive states $S_2...S_n$, then there is also something we can say about the direction in which the system will probably move. The interjection of the term *probably* is necessary given that we may not know everything about how the forces and states interact with one another, but the introduction of a probabilistic element does not obviate the fact that we can talk intelligently about the directionality of the system's motion. Of course a stable environment is a stipulation made for theoretical convenience, but on the other hand it seems that the environment does not change in a totally arbitrary (unpredictable) way. Nor, as we have seen already, is the environment independent of the organisms it affects. "It is not an exaggeration," writes Futuyma, citing Lewontin, "to say that by virtue of past evolution, species create their own environments....It is an error to think of species simply as passive sufferers of harsh external fate; they are active participants in a dialectical interchange between organism and environment" (1986: 19). Thus when a philosopher of science such as Dennett (1995) claims that the modern synthesis of evolutionary biology can offer a wholly mechanistic account of how life on the planet has reached its present state, he must not be read as denying a directional element. Whether directionality alone implies a teleology, however weak, will be seen presently.

At the same time that many evolutionary biologists and philosophers of biology embrace this mechanistic view, they also use what one might call the language of optimization to explain how the allegedly goalless process as a whole functions. This is true in the case of fitness. To simplify the account of how adaptation occurs, there is more than one mechanism by which an organism can achieve a certain degree of fitness in a given environment. Random walk is one, natural selection is another. Here again a direction and a goal of sorts emerge. For assuming that natural selection comes to dominate other factors which cause and preserve adaptation among organisms of type O, and assuming also that the environment E remains stable, then E determines a kind of *telos* for O-type organisms. Under the conditions specified (that natural selection is the strongest evolutionary force in E for O-type organisms and that E is stable), let us imagine that we take a Kuhnian view of the theory-ladenness and cultural dependence[50] of science and claim that the facts at hand can be accommodated by an Aristotelian (i.e., teleological) account of causes as well as by a mechanistic schema. That is, the force of natural selection which drives organisms of type O within environment E can be seen as an efficient cause, while the equilibrium (statistically normal) state itself can be viewed as a *telos* or final cause to which the O-type organisms tend.

We can then question whether the final cause (the stable equilibrium) "pulls" the evolution of the organisms or whether the equilibrium, understood as a *telos,* is simply a by-product of the efficient cause (natural selection) which mechanistically "pushes" the process forward. What would be the ground for deciding which view is correct? It might be thought that some sort of logical priority should inhere in the alternative which we would propose to be "the" pattern of evolution as a process. But no such priority is evident in this case. The mechanical efficient cause, natural selection, can be seen as the propeller of convergence to stable equilibrium, or in other words, the particular equilibrium state can be defined in terms of the environment, the O-type organisms' structure and behavior at a given point in the past, and natural selection as a force:

$$(1) \qquad O_{t=n} \xrightarrow{NS} O_{t=n+m}$$

where at time $t = n + m$ the O-type organisms are in a stable equilibrium determined by their stable environment E and by their previous state. In other words,

$$(2) \qquad O_{t=n+m} = f\,(O_{t=i,\ i<n+m}, E, NS).$$

Saying that the stable equilibrium is a function of the state of O-type organisms at some prior point, of the environment E, and of the force of natural selection is not to say that we can predict precisely how the state of stable equilibrium would look given these factors. This is because we presumably cannot take into account every relevant aspect of E and of the O-type organisms, nor do we know everything there is to know about how the force of natural selection operates. For instance, we do not understand every aspect of genetic combination. Furthermore, chance may play a role. Nonetheless, it appears that many evolutionary biologists and philosophers of science would feel comfortable with (1) and (2) above. Given what we know about the environment and about the state of the O-type organisms at some point in time, we can at least make an intelligent guess as to some aspects of the equilibrium state.

But suppose we know an early state of the O-type organisms, the environment E, and the state of stable equilibrium for the O-type organisms in E, that is, assume we know $O_{i<(t=n+m)}$, E, and $O_{t=n+m}$. Then presumably we would be justified in saying that we could begin to piece together what sort of force (again conceived as an efficient cause) brought about this state of affairs. In other words, analogous to (1) and (2) above, one could claim that

(3) $$O_{i<(t=n+m)} \xrightarrow{O_{t=n+m}} NS$$

or, correspondingly:

(4) $$NS = f\,(O_{i<(t=n+m),}\, O_{t=n+m,}\, E)$$

The point here is that if the equilibrium state and any prior state of O-type organisms within a stable environment are known, we can then say something about the force which, as efficient cause, leads from the organisms' prior state to their equilibrium state. In fact, that kind of reasoning characterizes much of the history of evolutionary biology: even before Darwin, Lamarck and others sought to describe a mechanism which would account for the organisms they observed in whatever environment they were to be found as well as for the fossil record.

Why does any of this matter? If an organism's Darwinian fitness is an intrinsic quality (e.g., a dispositional property, as Sober says), then the goal toward which that property drives the organism and others of its ilk *within a given environment* can be viewed as equally intrinsic. Moreover, one has the same warrant to dub this goal the motive force in the organism's longevity and production of offspring as one has to call the fitness-propensity the cause of these outcomes. For consider what it is to say that something is a force. A force in the sense used by evolutionary biologists is an abstract cause, that is, a cause which is recognizable in the abstract by virtue of its similar effects across a range of scenarios affecting a number of organisms. Thus we see organisms of a type experiencing varied life spans and reproductive successes. To the extent that these outcomes are not governed by chance, we infer the existence of a cause, or force, which brings them about. But the causal relationship is an association of events scattered across a time span. In general what we call the force behind a phenomenon such as evolution is temporally prior to its effects. This sense of temporal priority would seem to be violated if we named a *telos*, such as a state of stable equilibrium, the force which motivates the progression in question. Two questions need to be raised here.

First we should ask whether calling the state of stable equilibrium within a given environment the cause of a taxon's evolution by natural selection really violates the temporal priority of mechanistic cause over effect. After all, the claim would not be that the force involved comes into being only when the equilibrium state itself is reached. Rather, a pull toward the equilibrium state always exists as (again to use the phrase Sober 1984a employs) a dispositional property. Thus it would be merely a convenience to call the equilibrium state itself the final *cause*, without adding that associated with this final state toward which the evolution of the taxon tends there is also a force which is at work before the equilibrium is reached. We have no other name for this "pulling" force, and so as a linguistic artifice we choose to christen it according to the state to which it tends. Secondly we should question whether temporal priority is really a necessary quality of a proper cause. Some have seriously entertained the possibility that it is not (Dummett and Flew 1954, Downing 1958, Waterlow 1974).

Without attempting to answer the more challenging question of whether effects can precede their causes, it seems clear that effects and causes can come into existence *simultaneously*. This is apparently the case in contexts where gravity pulls two bodies toward one another. The force (gravity) and the effect of the force (the pull) bear no before-after relationship in our perception; they exist simultaneously. If this is accepted, at least provisionally, then perhaps the same simultaneity can be attributed to the case of natural selection and the stable equilibrium to which it would drive an O-type organism in a fixed environment: what many contemporary philosophers of science such as Dennett would regard as the effect -- the "pull" or "push" (depending on one's perspective) toward a stable equilibrium -- and what they would regard as the force which does the pushing or pulling -- natural selection -- come to exist simultaneously. In fact, depending upon the purpose at hand we can treat natural selection as being identical with this force. At other times we may wish to speak of natural selection as being in some sense the cause of the force in question rather than as being the force itself. Again the analogy with the force of gravity comes to mind. Upon seeing a falling object, we may exclaim that gravity is causing it to descend. By that we mean that gravity *is* the force which moves the object. Then again someone may say in a beginning class on astronomy, "Gravity exerts a weaker pull on a small, man-made satellite than it does on our moon." Here gravity is a sort of category which embraces forces of various magnitudes. Let us arbitrarily choose the word "pull" to describe the effect of natural selection in a given environment. Then what is this pull? At times we may consider it to be the force of natural selection itself, while at other times we may conceive of it as a particular movement associated with natural selection as a category of forces of various magnitudes. Yet if such motion is constant, that is, if it moves always in the direction of the stable equilibrium, then is it not associated with this final state as much as with natural selection? And if the association is as strong, what reason do we have for asserting that it is caused mechanistically by natural selection instead of teleologically by the goal-state of stable equilibrium to which the system tends? Can it be that we prefer the mechanistic push instead of the teleological pull arbitrarily (perhaps for cultural reasons, as one reading of Kuhn might suggest)?

At this point it might be objected that the association between the force and natural selection is much stronger than that between the force and the equilibrium state as *telos*. After all, the force can exist and move in *some* direction even though the environment may change, while because of the altered environment, the equilibrium state may prove to be much different than it would have been under the previous conditions. This objection, however, makes the mistake of identifying the force with natural selection, while almost in the same breath taking the force as an anonymous something and establishing natural selection as its source.

One way of handling such problems of interpretation is simply to avoid them. A means of doing so is suggested by the essentially recursive form of equation (2) above. By restricting the context of comparison to an immediate one -- one state of an organism as compared to the immediately previous state -- we can "move through" a progression of states without having to determine whether the progression as a whole is mechanistically or teleologically describable (or both). This is not to say that the general question of the pattern's nature is not of interest. On the contrary, it is a fascinating problem and the search for a solution may be rewarding. But there is no *need* to solve it in the context of a recursive model of evolution as a whole (including a recursive model of fitness).

## (4) Design without a designer

It is perhaps an accident of history that evolution -- the mere word -- evokes not just a list of positive theses, but also a negative comparison with outlooks which allegedly stand in fundamental opposition to the "key" aspects of evolution. Two such sterotyped perspectives, usually perceived as being closely related to one another, are creationism and teleology. At the risk of oversimplifying, in the minds of many evolutionists the threat of creationism amounts not just to the imputation of design, but more importantly, of design undertaken by a purposive designer. This last qualification is important because for many students and practitioners of evolutionary biology's folkways, the presence of such a designer would directly oppose the dual mechanisms of evolution -- chance (in the form of genetic drift, for instance) and natural selection. But it should be emphasized that one can assert design, even in a teleological sense, without implying the existence of a designer. In fact the word "design" appears often enough in the literature of evolutionary theory in a neutral sense. An organism's *Bauplan*, for instance, constitutes a design in a sense which would generally be thought consistent with purely mechanistic explanations of life's origins and progress (Gould and Lewontin 1979). For many the difference between the two senses of design seems to be this: "design" consistent with mechanistic evolution describes physical structure without further implication; "design" as the vision of a creative agent implies a final state. This distinction makes clear why a link is often drawn between creationism and teleology. A mechanistic account of how life has emerged and developed does not appeal to a design *qua* perfection of the developmental process.

At least that is a standard account of the opposition between evolution on the one hand and teleological creationism on the other. It has become a truism that there is a basic opposition between the outlooks, even though some influential voices have tried to reconcile the theses of evolution with a kind of teleology (notably Teilhard de Chardin 1956, 1957). Because of the still-unfolding history of the debate between creationists and evolutionists, a debate marked by high emotion and frequent distortion of the opposing position, it would be well to view the basic matter without appeal to the old terminology. The terms are loaded and are liable to distract us from important truths.

Instead of using these terms, we might try to characterize evolution in terms of the concepts of "pushing" and "pulling" forces discussed in the preceding section. We will consider both pushes and pulls to be forces or causes, that is, things that can motivate change. Now when we say that the environment favors a certain phenotypic quality, in the context of evolution we are claiming that organisms possessing that phenotype will (*ceteris paribus*) live longer than otherwise similar organisms which do not have the same structural or behavioral characteristic. This notion of the environment favoring a certain quality has the earmarks of a force or a cause: favoring or disfavoring affects the number of offspring which will be born in the next generation. (There are other measures of the effect, of course. Dawkins' notion of the selfish gene would have us pay attention to the number of genes of a certain kind which appear in the next generation.) Alone, the purely temporal aspect of the situation does not give us much help in determining whether the force in question is a push or a pull. One generally thinks of a force as preceding or as being simultaneous with its effect (as in the case of gravity). Whatever the temporal fact of the matter happens to be, the notion of an environment favoring a certain phenotype can be reduced to talk of pushing or pulling. The favorable phenotype can be conceived as pushing the organisms which have it toward greater success in the environment (forward-looking perspective on the force) or as pulling those which had it toward their present, thriving status in the environment (backward-looking perspective). Talk of forward and backward perspectives belongs more to a spatial metaphor than a temporal one, but these spatial aspects likewise do not grant privileged status to either notion, pushing or pulling.

To take the case even further, it seems to be true that for a given environment, one might imagine an optimally adapted organism. Perhaps such a conception is possible only for very simple cases, since the synergistic interactions of all an organism's properties are presumably too complex to conceive. (This is consistent with Rosenberg (1985), who argues that theoretically the principles of biology are the same as or *reducible* to the principles of physics. But practically, Rosenberg continues, reduction of biology to physics and chemistry is impossible.) Moreover, in accordance with Sewall Wright's concept of adaptive landscapes, perhaps only a "local optimum" -- the top of a foothill, so to speak -- is achievable (Dawkins 1982: 43 - 47; Ruse 1995). And it may also be the case that a number of possible structures are of roughly equal adaptive value. So let us consider a *very* simple case. If there is an organism which is relatively small and soft-bodied, whose coloration functions as camouflage under some conditions but not under all, and which lives among many carnivorous species, such an organism will need some extremely keen sense to warn it of approaching predators and some structural feature to allow it to escape. In such a scenario, both pushing and pulling descriptions seem appropriate. One can say that the environmental factors *push* the organism in question to develop a keen sense of hearing or smell or sight or some combination of these, along with the long legs or wings or something which allows for sudden, rapid, and prolonged locomotion. On the other hand, one could just as easily analyze the environmental factors, establish a model organism in accordance with the environmental analysis, and then argue that small, soft-bodied organisms in the environment are developmentally *pulled* in the direction of the model survivor.

By this point it should be clear that what we have just called a pulling explanation of evolutionary development is quite teleological, but such a teleological model need be no less mechanistic than a "pushing" explanation. In other words, there is no necessary relationship between explanations of development and evolution which are teleological in character and those which posit the existence of an intelligent designer or self-aware agent. That this link is often drawn is perhaps an accident of history; in any case it is not necessary. This makes good common sense when one stops to think that within a given environment there is presumably an optimally designed organism or set of organisms, even if only in theory. To the extent that an observer can conceive of this design, a teleological yet mechanistic explanation of evolution can be offered.

## 4. Speciation

There are two major problems involved in reconciling divisions of life with the fossil record. Besides the fact that the fossil record is sometimes too sparse to serve as an adequate "check" for model taxonomies, there is also the problem that we have not perfectly reconciled concepts which, intuitively speaking, seem to be important to to the taxonomic exercise. Complexity is a case in point.

> There is little logical order in time of appearance [of phyla in the fossil record]. The Arthropoda appear in the record as early as do undoubted Protozoa, although by

> general consensus the Protozoa are the most primitive phylum and the Arthropoda the most 'advanced' -- that is, structurally the most complicated -- among the nonchordates.... (Gaylord Simpson 1967, p. 33; see also pp. 34 and 35).

The same problems affect our appraisal of other phenomena associated with evolution, e.g., extinction.

### (1) Extinction

Based on what was said above about the concept of a species, it should be no surprise that the notion of extinction -- regardless of how simple it might appear at first glance -- is actually problematic. The fossil record displays plants and animals which have ceased to live in two senses. The obvious sense applies to the individual and is not terribly instructive: the individual represented by a given fossil is of course no longer alive. But we are tempted to say that the *type* of animal represented by a specific fossil has likewise disappeared, or in other words that it is extinct. The problem with such an assertion, as we have seen above, is that it is not clear what is meant by "type." Paleontologist Simpson categorizes animal life by asserting the existence of

> a relatively small number of basic themes, general types of organization. The broadest of these themes are those formalized by zoologists as phyla, each of which represents a distinct plan or level of anatomical organization, with its attendant possibilities in the functioning of the animals. (1967 p. 25)

At this broad-brush level of observation, it is not just structure which is important, but something a bit more mysterious, namely, the "level" of structure. More will be said below about this difference between structure and the interpretations of structure which are given names such as "level of anatomical organization" in Simpson's phrase.

What is it, then, to which we apply the predicate "is extinct"? It is not the individual. Individuals die, but properly speaking they do not become extinct. But surprisingly, the broadest organizational scheme imposed upon the record of life by taxonomists also does not admit of extinction. "In spite of possible exceptions involved in the largely verbal question of defining 'phylum,' it remains true that *no* *major basic type of animal organization is known ever to have become extinct* (Simpson 1967, p. 39, n. 5; Simpson's emphasis).

This fact can be interpreted in two ways. First, it may be taken simply as a feature of the history of life, in which case it can be explained as a coincidence or as the necessary entailment of some rule. It might be claimed, for instance, that so long as a pattern of organization fills an ecological niche, there is only one reason for organisms constructed in accordance with that pattern to become extinct. That reason is the existence of competitors, but at the level of the phylum, which paleontologists and others describe with such terms as "basic," the broad structures tend to be mutually exclusive. In other words, finer distinctions among organisms -- classes, genera and species, for instance -- necessarily share some characteristics (the ones which unite them under the umbrella of the higher levels of classification) and these similarities mean an overlap of ecological niche can occur in the "real world." Notice that these lower levels are established on the basis of similarity and difference, whereas the higher levels are established to emphasize raw difference. Or the observation that no phylum has gone extinct might be taken as evidence that there is a certain analytic element in the concept of extinction. Perhaps it is more than just the vagaries of natural forces that have kept any phylum from going extinct. Maybe we define phyla in such a way that they cannot become extinct.

There is an interesting connection here with the notion of a *Bauplan*, especially as the concept is used by Gould and Lewontin (1979). (The concept was originally developed by the nineteenth-century German *Naturphilosophen* and according to Ruse (1995: 92 - 93) was perpetuated in American evolutionary thought through the influence of Spencer, Henderson, Wright, and their colleagues.) As we saw in the last chapter, Gould and Lewontin's observation constituted an attack on more purely adaptationist views of evolution. Dennett, for instance, rails against Gould and Lewontin for departing from the view that changes in the structures of organisms -- which means evolution itself -- is driven primarily by natural selection (1995). There is no need to repeat all the particulars of the debate. To review what was said earlier, Gould and Lewontin's basic position is that many aspects of organisms are forced by some higher-level feature of structure. Metaphorically, a cathedral's arch forces the presence of a spandrel, so that it would be incorrect to say that the spandrel, undeniably visible though it is, was in any sense designed into the structure. This view ties in well with the fact that Gould in particular has championed the notion of punctuated equilibria rather than espousing continuous, gradual change (Gould and Eldredge 1977). If many features of organisms are forced by the overall *Bauplan*, then one would not expect many structural changes except in cases where the entire *Bauplan* varies. Such large-scale change, in turn, is unlikely to occur continually, and thus high-level groupings such as phyla are liable to continue to exist rather than become extinct.

It is worth reflecting with some care what such a statement means within the scheme of evolution as a whole. Suppose we begin by proposing that organisms can be grouped, conceptually, on the basis of structural patterns and (to modify Simpson's phrase somewhat) the attendant possibilities of function. In other words, we distill morphology, at least in part, by imagining what function certain structures might fulfill. Armed with this division of life into groups, we assert that certain groups have become extinct. This has occurred for two reasons: pure chance, which frustrates any search for a pattern, or ill-adaptedness to the environment. This latter reason clearly has to do with function, so that one may suspect there is no wonder that the broadest of our functionally defined levels of organization has suffered no extinctions.

### (2) Goals of a theory of speciation

Like most of the parts of science as a whole, the theory of evolution seeks to explain the phenomena we see, including the fossil record and the existing diversity of life. But what we call "explanation" in science is generally an attribution of cause: "Such and such a phenomenon *is the result of, is caused by*, something else." One need not have studied Aristotle to realize that causes can fall into more than one basic category, nor need one have read Kuhn to believe that the process of attributing causes is influenced by culture. The theory of evolution is held by many to be patently mechanistic rather than teleological in character, at least at its best. That means that it seeks to explain the variety of life and the fossil record by describing causal processes which "push" rather than "pull." "Pulling" is what teleological accounts, or in other words, final causation, would rely upon.

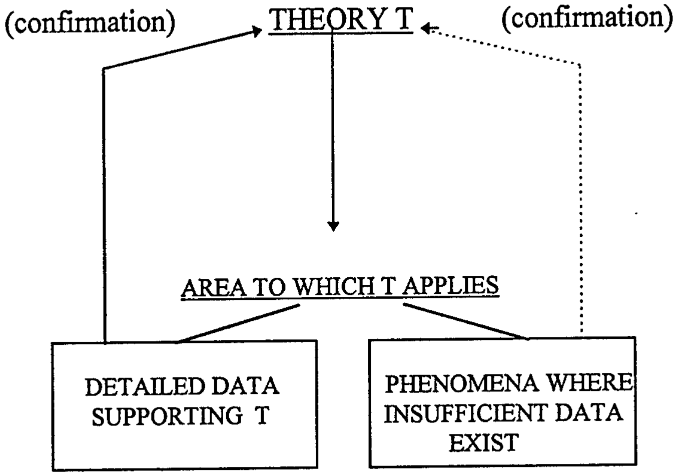
But it is very difficult to describe a process without specifying a product, and the product can in turn be interpreted as a kind of goal of the process. If there is such a goal in the case of Darwinian evolution, it is the *fit* (adapted) organism. But the contemporary evolutionist, enamored of the ideal of mechanistic causation as she is, will quickly point out that fitness is not an absolute term and is therefore not a goal. Fitness is in fact dependent on the specific selective environment. Moreover, some accounts hold that at best we can approximately describe fitness (e.g., through a propensity interpretation).

The theory of evolution begins its enumeration of mechanistic causes by suggesting how "simple" life may have emerged mechanistically from non-living matter, and then elucidating the processes by which the heritable structures of organisms change. The program as a whole is therefore a bit like a game of "connect the dots": the goal is to create an intelligible picture of how life on the planet emerged and developed. What is more, the picture is supposed to be more persuasive than rival pictures. How the development of life on the planet actually proceeded has been a mystery throughout most of human history. Myth was the only means of approaching the subject. ("Myth" is meant here in a broad, non-pejorative sense as a non-scientific account. What constitutes "science" is of course a problem in its own right, but for the moment I rely on the reader's intuition.) The specific details of the development of life -- exactly how every species between *Australopithecus* and *Homo sapiens* looked and acted, for instance -- remain largely a mystery to this day, but at least we have a *sketch* depicting how life emerged and evolved. Part of the sketch is fact and part is hypothesis, but (arguably) the picture makes sense despite the absence of so many details. The facts or "dots" available to us -- the data points offered by the fossil record and by observations of existing species -- can be connected based on hypotheses derived from allied branches of science. Genetics, for instance, explains how morphological change can originate from mutations among genes and how such change can be propagated among the generations of a population in statistically predictable ways. Geological theories of shifting tectonic plates allow us to envision our planet as it existed millennia ago, when continents now separated by oceans which are impassable to many land-dwelling species were connected with one another. Evolutionary biologists thus have conceptual tools with which to explain how physically isolated areas of the planet in its present state -- islands, for instance -- might have been populated by species now extinct on the mainland. Physicists' insights into the decay of radioisotopes allow theorists to establish chronologies more precisely than was possible merely by looking at geological strata. With the aid of such subtheories borrowed from genetics, geology, physics, and other sciences, evolution builds hypotheses as to how life emerged, how speciation occurred, and how changes in structure have been propagated across generations of a given species.

But for all the connections between evolutionary theory and other sciences, the gaps between facts -- the lines between dots, as it were -- are bridged on a faith made of inference. The lines connecting the dots are drawn without anyone having seen the actual stages of development which the lines represent. That is not to say that to draw the lines is an arbitrary act; on the contrary, the lines represent what the majority of the scientific community finds to be most plausible. Now this plausibility has a momentum of its own. In the absence of data points linking two very similar species representing different epochs in the fossil record, let us say, many evolutionary biologists assume that there must have been a number of gradual morphological changes which led from one species to another. Punctuated equilibrium fans offer a somewhat different sketch of evolution's progress. In either case, one brings one's assumptions to the table. Of course there is nothing particularly unusual about this facet of analysis. Science proceeds largely by applying hypotheses which are presumed to be correct until they are demonstrated, through failed application, to be false. The process proceeds something like this: based on a background of data, subtheories, and perhaps general culture, a "macrotheory" is established. Following the general acceptance of this theory, pieces of the theory's "bailiwick" are assumed to conform to the general principle, even though data is lacking. So long as the theory explains most of the data in its realm of application better than rival theories, it is assumed to be correct even though data which would prove its aptness are lacking in some areas. In a backhanded way, such areas of scarce data can even be taken as evidence in favor of the theory so long as no rival theory seems to do a better job of explaining the data in general. This phenomenon is represented by the dotted line in the diagram below.

As the arrows in the diagram suggest, there is a sort of circularity involved here, though certainly not a harmful one. The theory *applies downward*, one might say, while the data *confirm upward.* That expresses not just the direction of "movement," but also the order: first the theory is applied, then it is confirmed. "Circularity" when applied to anything resembling an argument or an assertion has a negative connotation, but to repeat, what we could see as a kind of circularity in this case seems to be perfectly acceptable. In order to set this sort of circularity apart from other types, which may in fact be something we should avoid, it seems wise to use something more than the mere word "circular" to describe the rhetorical movement in this kind of scientific theorizing. "Upward circularity" seems appropriate, since we are emphasizing that the last part of the movement ends at the level of theory. (Of course this term is also a bit arbitrary, since we could redraw the diagram to show "THEORY T" at the bottom of the picture just as easily at the top, but nonetheless the name will suffice to distinguish this type of circularity from other sorts.)

Upward Circularity


(confirmation)
THEORY T
(confirmation)
AREA TO WHICH T APPLIES
DETAILED DATA SUPPORTING T
PHENOMENA WHERE INSUFFICIENT DATA EXIST

Now if there is such a thing as "upward circularity," it is natural to ask whether there might also be a kind of reasoning employed in evolutionary biology which moves in the opposite direction. The concept of upward circularity as defined above emphasizes the confirmation of theory: a theory is applied to data and, if the application is acceptable -- meaning it proceeds in some sense more smoothly than the application of any rival theory -- then the current theory is in some measure confirmed. But what if our focus were not on the theory but rather on particular data? Explaining the phenomena we encounter is, after all, the goal of theorizing. "Being explained" rather than "being confirmed" is the goal for data. In this regard, let us consider an important subtheory within the overall theory of evolution.

The Hardy-Weinberg theorem was posited independently in 1908 by G.H. Hardy and W. Weinberg to explain why a numerically predominant allele does not eventually become the only allele at a locus in a given population in the absence of any forces other than the statistically predictable inheritance principles discovered by Mendel. One might have thought that the allele which was dominant in the first generation would become even more ubiquitous in the next, and that this increasing dominance would continue until the allele which was least frequent in the first generation would cease to exist in the population. This line of reasoning assumes that no other forces which might favor the less common allele -- chance or natural selection, for instance -- come into play. The Hardy-Weinberg theorem shows that even (or especially) in the absence of such forces, the relative proportions of alleles remains constant, or to use other language, that populations remain in "Hardy-Weinberg equilibrium."

Consider a numerical demonstration of the theorem (following Futuyma 1986, pp. 82 ff.). For simplicity's sake, we focus on two alleles at a locus, A and A', so that there are three genotypes, AA, AA', and A'A'. The frequencies of these genotypes are respectively denoted as D, H, and R. D is the ratio of the number of AA genotypes, denoted nAA, to the total number of individuals in the population, N. In other words, D = nAA/N. Similarly, H = nAA'/N and R = nA'A'/N. These frequencies are conceptually normalized so that D + H + R = 1. Obviously there are two copies of A in every AA genotype and one copy in each genotype AA', or 2D + H copies of A. The frequency of A, called the allele or gene frequency, is the ratio of the number of copies of A, 2nAA+ nAA', to the total number of genes at that locus in the population, which is 2N. (Remember that every member of the population has *two* genes at the locus under consideration, so that the ratio expressing gene frequency has 2N rather than N as the denominator.) Thus the gene frequency of A, denoted p, is (2nAA + nAA')/2N = nAA/N + nAA'/2N = D + H/2. The frequency of A' is denoted q. If we normalize again, so that p + q = 1, then q = 1 - p. Following the same pattern of reasoning as for the frequency of A, we can also say that q = R + H/2. Then it can be shown that for any generation, p and q remain stable. Futuyma reports the results of a study (done by Ford, 1971[51]; Futuyma 1986: 83) of 1612 tiger moths (*Panaxia dominula*), of which 1469 had coloration associated with the genotype (AA), 138 were colored consistent with (AA') and 5 were (A'A'). The results were clear-cut:

(1) (Frequency of A) = p = (1469/1612) + (1/2)(138/1612) = 0.954.

(2) (Frequency of A') = q = 1 - p = 0.046.

The frequencies of the genotypes were then:

(3) (Observed frequency of AA) = D = $p^2 = (0.954)^2 = 0.9101$

(4) (Observed frequency of AA') = H = $2pq = (0.954)(0.046) = 0.0878$

(5) (Observed frequency of A'A') = R = $q^2 = (0.046)^2 = 0.0021$

The expected genotype frequencies if the organisms were in Hardy-Weinberg equilibrium would be:

(6) (Expected frequency of AA) = (0.9101)(1612) = 1467.0812

(7) (Expected frequency of AA') = (0.0878)(1612) = 141.5336

(8) (Expected frequency of A'A') = (0.0021)(1612) = 3.3852.

Futuyma considers the observed and expected frequencies to be close enough to demonstrate that the locus of moths was in Hardy-Weinberg equilibrium.

Leaving aside the details, the pattern of the argument goes like this:

(1) Abstract expectations are formed for the frequency of alleles and genotypes among diploid, sexually-reproducing organisms under certain idealized conditions.

(2) The frequency of genotypes is empirically observed for a given population.

(3) The abstract expectations are compared to those observed and a statement of the relationship is formulated.

The goal of the argument is simply to say that the idealized conditions probably applied in the previous generation of the organisms under consideration. In other words, the assumptions of Hardy-Weinberg equilibrium are borne out in the present generation. Notice that the hypothesis of equilibrium is directly relevant to the fitness of the organisms in the previous generation. Had the interaction of genotypes and selective forces been other than neutral, we would expect to see some deviation from the Hardy-Weinberg expectations. That we did not see such divergence is *prima facie* evidence that the fitness of the alleles did not undergo a change in the previous generation.

This pattern of reasoning jibes well with a recursive understanding of fitness, one which considers the fitness of a level of life in terms of the fitness of the previous level. Supposing we had a list of similar observations of gene frequencies corresponding to various generations, we would be able to generalize about the respective fitnesses of a generation's genotypes by viewing the proportions of those genotypes in any generation as a function of previous generations' genotypical distribution.

Evolutionary biology differs from many fields of scientific analysis in that prediction of future occurrences is almost entirely excluded from its realm of concern. Thus Futuyma calls evolution a "historical science" (1983: 226). This is true even though many related sciences use the insights of evolutionists to predict what will happen to various phenomena -- a zoologist may predict the decline of a long-isolated island-dwelling species once new predators are introduced, for instance -- but evolution concerns itself primarily with the past and present.[52]

The historical and contemporary phenomena evolution seeks to explain can be divided into two broad categories, corresponding to past and present. First, there is the fossil record, which displays numerous extinctions as well as new additions of life forms into the world. The span covered by this record of life is enormous -- upwards of three billion years by most accounts. Second, evolutionists interest themselves in the origin of existing species and related phenomena. These species show a sometimes bewildering array of similarities and differences in structure and behavior. As we have seen, it is not clear that there is a privileged means of grouping existing organisms based on common structure, nor can they be grouped once and for all based on behavioral phenomena such as reproduction.

Inextricably bound to structure as it is used in the explanations of evolutionists is the notion of *function*. Evolutionists take it as evidence in favor of common ancestry, for instance, when they observe that a molecule in the blood of one organism performs the same function as the same sort of molecule in the blood of a very different organism. One can also see a close relation between function and *behavior* as evolutionary biologists use the terms. Put simply, organisms viewed in isolation *behave*, or they *function* within an ecological system, while parts of organisms simply *function*. This is of course a matter of linguistic convention. It is perhaps more than coincidence that the best know articulator of the important relationship between form and function was a biologist as well as a philosopher. Aristotle tells us that the most important clue to a thing's proper function is its form. In evolutionary biology, however, this aspect of form -- its ability to tell us about function -- works only when we can relate parts to one another.[53]

Arguably we can even witness the emergence of new species under laboratory conditions, which is naturally of enormous interest to evolutionists. Constructing a theoretical framework to account for all of the past and present evidence of how life has emerged and changed is the goal of the theory of evolution. The first assertion one generally encounters when reading books about Darwinian evolution is that two theses are central to the overall theory: first, that all currently existing species derive from a common ancestor and second, that natural selection is the primary mechanism driving the course of evolution. Let us briefly examine the first of these two claims.

### (3) Descent from a common ancestor

Implicit in the assertion that all life is descended from a common ancestor is the claim that this common "mother" arose mechanistically from non-living matter. How? Various detailed hypotheses have been put forth, but the basic thrust of all these versions is that complex molecules developed the ability to reproduce themselves or parts of themselves. Laboratory experiments have managed to produce such molecules, but to date no one has been able to create life from non-living matter. Nor is there hard evidence that life was created in this way. As Gaylord Simpson puts it, "there is no known *record* bearing immediately on the origin of life (1967: 15; my italics). If all life has evolved from a common ancestor, then presumably a "map" of evolution could look something like the familiar tree diagrams. Clearly the theory requires a subtheory -- one that accounts for the "branching" which represents speciation. Evolutionists rely on genetic mutation to explain how structural changes arise in the first place. In so doing the theory tacitly rejects so-called Lamarckian evolution, which asserts that behaviorally or environmentally acquired characteristics -- the giraffe's long neck acquired by stretching is one paradigm -- are the jumping-off point for evolutionary change. A further challenge is to explain why an organism of a sexually-reproducing species is not exactly like either one of its parents and why it is also not a perfect blending of the characteristics of both parents. Mendelian inheritance offers an escape from the possibility of identity with one parent as well as from blending inheritance, so that the offspring can be unlike either parent and at the same time need not display an "average" of the parents' characteristics.

### 4) Natural selection as the primary agent of evolution

The second key theory of Darwinian evolution -- that natural selection is the primary agent of evolution -- can also be problematic. The basic idea behind this assertion is that certain organisms will live longer and produce more offspring than their peers because their particular structures and patterns of behavior afford them corresponding survival and reproductive advantages in a given selective environment. To the extent that these differences in morphology and behavior are genetically based, they are also heritable. Thus, *ceteris paribus*, not just particular organisms but also certain lineages will tend to survive while others will tend to die out. Yet it is not clear what contributes to a "superior" lineage. At first glance, it would appear that those organisms which live the longest and produce the most offspring would tend to engender such lineages, but observations of organisms do not always support this thesis, or at least not in a straight-forward way. As we have seen, Morris (1986) found that litter size among white-footed mice does not correspond in any simple way with the number of offspring which will reach reproductive age. And the most frequent litter size (four) in Morris's experiment was not the most productive. The largest litter sizes were the least productive. However, "[t]here were no obvious physiological reasons for the poor recruitment of young from large litters" (Morris 1986: 173). Clearly certain assumptions would have to be made in order to view these data as being consistent with a Darwinian model of evolution. For instance, one might assume that (1) chance distorted the results carefully gathered by Morris; (2) litter size is somehow converging toward the optimum size (five) but has not yet reached that point; or (3) it is somehow more advantageous to contribute fewer individuals who live to reproductive age to the next generation.

The theory of genetics itself, even at the level of basic distinctions among types of cells (e.g., somatic versus germ cells), seems to track only imperfectly with observed results. After detailing their findings that acquired tolerance to antigens can be transmitted to first- and second-generation offspring in mice, Gorcynski and Steele assert that this phenomenon "severely challenges Weismann's doctrine [that characteristics of somatic cells do not influence germ cells], and seems to us to demand an explanation not offered by conventional neo-Darwinian genetics" (1981:

680). By this they apparently mean that they have observed a kind of evolution which is Lamarckian rather than Darwinian in so far as environmentally-induced phenotypic change was propagated. In their research it was clear that not just the somatic cells of the parent organisms had been directly affected by the environment; germ cells had been, too. This underscores one of the hallmarks of the Darwinian theory of evolution as it has been constructed in the so-called new synthesis. By this account, genetic mutation as the result of *random* "copying errors" is the beginning of the process of evolutionary change. In this context, *random* means that there is no clear link to environmental features. Hill implies as much when he introduces his discovery of environmentally-induced heritable changes in tobacco plants. "A further distinction arises in so far as chromosomal mutation is a random process, whereas conditioning, like many induced extrachromosomal mutations, is a non-random process" (1967: 735). The problem is that Hill, like Gorcynski and Steele, observed a kind of inheritance which appears to be Lamarckian.

Recursive fitness cannot rescue evolutionary genetics from these problems. But importantly, a recursive concept of fitness is not as liable to distortion by apparently Lamarckian phenomena as other concepts of fitness may be. If we think of fitness as being a propensity to a certain degree of reproductive success, for instance, and if we simultaneously declare the unit of selection to be the gene, then we must also conclude that the genetic status quo of an organism *determines* its fitness. If environmental factors suddenly impinge in Lamarckian fashion, so that acquired characteristics are perpetuated in future generations, our fitness expectations may be confounded. By contrast, a recursive concept of fitness need not bridge from the fitness expectation at one moment, before environmental factors have had an effect on an individuals germ cells, all the way to another moment when the reproductive expectation of the organism has been changed by environmental factors. Instead, the transition can be made so gradual that it is effectively seamless, without violating our formal definition of fitness:

$$\text{fitness}_{t=i} = f\,(\text{fitness}_{t=m}),\ m < i.$$

## 5. Evolution's "causal boundaries"

The modern theory of evolution has two or three (depending on one's theoretical commitments) of what I will call "causal boundaries." The meaning of this term will be made clear shortly.

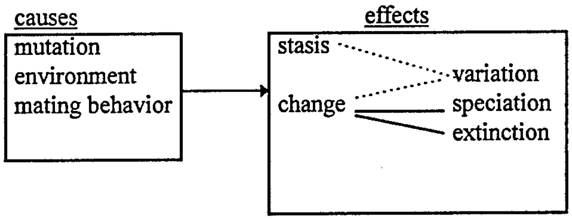
The first causal boundary is genetic mutation, which explains how heritable changes in organisms emerge. To explain how the various phenomena in the fossil record and in contemporary observations jibe with one another -- to connect the dots and make a complete and understandable picture, so to speak -- the proponent of evolution must have a means of explaining how morphological changes arise. It is all the more convenient if this basic source of change can help account for the propagation of changes across generations of a species. Genetic mutation accomplishes both tasks to some degree. That genetic mutation sometimes results in phenotypic change can be demonstrated in the laboratory and inferred from homelier breeding activities, such as in the dog and horse breeding industries.

A major point of contention among evolutionists is the magnitude of mutation. Classical evolutionary theory and the new synthesis regard only *gradual* change as efficacious in evolution. Genetic leaps do occur by this account, but such macromutations tend to be fatal to the altered organism and thus radical change can only occur as the sum of tiny, incremental changes across long time spans. It is not possible to offer a rigorous definition of "long" in this context; suffice it to say that the breed of evolutionists who favor gradualism (call them gradualists) oppose saltationist or "leaping" models of how speciation has occurred and can occur.

The second causal boundary is the environment, which evolutionists assert affects the ways organisms change and the duration and nature of their periods of stasis. In an ontogenetic context the claim seems obviously true: we can observe how some organisms are better suited to their environments than others. The classical view of the environment is thus what we can call a "pruning force": the environment destroys those organisms whose morphology or behavior is second-rate compared with organisms which are at least well enough adapted to reach maturity and reproduce. As we have already seen, a complicating factor is introduced by observations which indicate that to some small degree Lamarckian evolution does in fact occur. Durrant (1962) and Hill (1967), for example, found that fertilizer-induced changes in tobacco plants were in fact propagated in future generations. A similar phenomenon has been observed in flax plants (Cullis 1983; these results as well as those of Durrant and Hill are discussed in Futuyma 1986: 44-45[54]) .

The third causal boundary is mating behavior. This boundary may be unnecessary to a mechanistic theory of evolution in that the first two -- genetic mutation and environment -- suffice to show how the process of evolution could have created existing species as well as the fossil record currently available to us. But this third boundary is particularly important in some theories of speciation and thus merits special attention.

The method of arguing for evolution can be portrayed in the following way:



It is important to note that evolutionists recognize constraints on the degree of variation within a given clade. It has been speculated, for instance, that despite evidence of rapid growth in intelligence among certain primates related to human beings, we as a species may be at a point where a decrease in intelligence is as likely as an increase. One way of characterizing the mechanism which controls this development is by drawing a distinction between passive and driven developmental processes.

> The significance of the passive-driven distinction lies in the independence among hierarchical levels implied by the passive mechanism, a counterintuitive notion for many. If this independence is overlooked, it is easy to think that large-scale behavior must be nothing but small-scale behavior amplified and to make unjustified inferences from large to small scale. The reverse inference is equally facile. For example, Jastrow (1981) discovered a trend in primates toward greater intelligence and inferred that, if past forces continue to operate, intelligence is likely to increase in the human lineage. If such a trend in primates exists and if it is driven, that is, if the trend is a direct result of concerted forces acting on most lineages across the intelligence spectrum, then the inference is justified. But if it is passive, that is, if forces act only on lineages at the low intelligence end, then most lineages will have

> no increasing tendency. In that case, most primate species -- especially those out on the right tail of the distribution like ours -- would be just as likely to lose intelligence as to gain it in subsequent evolution (if they change at all). More generally, inference from large to small scale, or vice versa, is warranted for driven systems but not for passive. Thus, to justify such inferences, whether they are used in prediction or, as is more common, retrodiction (historical analysis), knowledge of the trend mechanism is essential. (McShea 1994: 1761)

Though couched in the circumspect language of the veteran analyst and researcher, this statement is rather shocking. For consider what it means in a wider context. Depending on how much of the evolutionary tree is passive rather than driven, evolution as an unbounded linear process could cease entirely, or at least within certain clades. Any "evolution" occurring in such a context would be merely cyclical.

It is tantamount to beating a dead horse, but it must be said anyhow: the passive-driven distinction has no appreciable effect on a recursive understanding of fitness. Should a species "devolve" in some respect such as intelligence, the fitness of any individual as well as that of the species as a whole remains consistent with the formal definition: a less intelligent stage can be a function of a more intelligent one as well as vice versa, and it does not matter in what order they come. The same cannot be said with certainty in cases where fitness, to whatever degree it is reckoned to depend on a passive property, is seen as a static propensity. That is because the "curve" -- the tendency -- may change direction unpredictably, thus defeating the original purpose of constructing a propensity interpretation rather than one based on actual numbers of offspring produced.

# Chapter Eleven: Conclusions (The Generality of Recursive Fitness)

In this chapter we explore the possibility of applying the concept of recursive fitness to a debate in historical phonology. The goal motivating the attempt is to test the recursive paradigm against an example of evolution which is not embedded in evolutionary biology. Here the distinction between the biological and non-biological is to be understood quite narrowly (and arbitrarily). What is not taken as belonging to the realm of evolutionary biology in most general treatments of the subject (e.g., a textbook such as Futuyma 1986) is our target area. If the recursive understanding of fitness is valid in the biological realm, it is so only because it makes sense to think of a certain class of concepts with the aid of a dynamic rather than a static paradigm. One might describe this class as something like the set of concepts related to dynamic processes. Heuristically, if something is in motion we need a technique or tool for atomizing the motion before we can adequately describe or analyze it. (Recall the analogy comparing recursion and calculus above.)

But surely other "historical" subjects also interest themselves in dynamic processes which occured as they did because the alternatives were somehow less appropriate -- less *fit*, metaphorically speaking. In this sense it is probably clear in a *prima facie* way that many disciplines outside of the conventionally defined realm of biology can be viewed as studying evolving phenomena. Economists look at certain institutions and philosophies as having what amount to life cycles -- emerging, maturing, and dying out, while giving rise to new but related phenomena of the same basic type. The same could be said of historians and indeed of any historical discipline. "Movements" in the realms of history, literature and art are often described in much the same language as evolutionary biologists define taxa: something is better adapted to a particular environment (only the term is often taken to refer more to a cultural or social milieu than to the natural surroundings) and therefore supplants predecessor movements which can nonetheless be seen as its "genetic" ancestors (in so far as elements are carried over into the "progeny"). Although human agency and design obviously play a role in human history generally, movements in various realms are frequently conceived in mechanistic terms, that is, as leading their own evolving "lives" which transcend the individual and even the collective plans of the human agents involved. To put the matter in a nutshell, one goal of historical inquiry in general is to reconstruct the ways in which various entities have evolved.

> ... [E]volution is a unique process, in principle not repeatable, which defies direct analysis and allows only reconstruction. In that regard, incidentally, the biologist is not alone: All historical disciplines, such as cosmology, archaeology, or Mediaeval heraldry, have the same problem. (Haszprunar 1994: 131; my trans.[55]).

Such speculations are rather vague and hard to pin down in such a way that we can test the generality of recursive fitness. What we need is a discipline where very specific claims are made about well-delineated, evolving phenomena, and preferably a discipline which employs at least some of the terminology of evolutionary biology. Fortunately such a discipline appears to exist, as evidenced by the particularly close relationship between historical linguistics and evolutionary biology proper. Perhaps the similarities arise because the general outlooks and vocabularies of both fields have become a part of everyday life. Polemical works written by contemporary evolutionists -- generally against the claims of so-called creationists -- appeal to historical linguistics as a realm in which the mechanistic development of new languages from old is widely acknowledged to have taken place. The rhetorical tactic is clear: first find a discipline which incorporates arguments sharing some of the abstract elements of Darwinian evolution but which is generally not the forum for emotionally charged debate of the sort that rages between creationists and evolutionists; next show that the concrete evolutionary claims in that discipline are credible; then argue that if these specific assertions are believable, the abstract pattern may be plausible as well; finally show that Darwinian evolution in the conventionally biological realm is nothing more than a conservative application of the credible pattern of reasoning. Historical linguistics seems to be especially beloved as a "straight guy" of this sort (e.g., in Kitcher 1982; Futuyma 1983).

On the other side of the coin, historical linguists find a treasure trove in a field where headline controversies makes even the layperson somewhat conversant with terms and arguments. A typical passage from a contemporary work demonstrates the way a linguist might use evolutionary biology to set arguments native to his own field into sharper relief.

> It [language] is in fact a structure that is some ways resembles a living organism more than a cultural artifact. Languages of course aren't really organisms, nor are they like them in any very deep sense; but because they're both historically evolved objects, they have enough in common so that certain biological modes of description can at times -- as analogies anyhow -- be quite appropriate. Both languages and organisms are systems, not just collections of parts. Or better, systems of systems, all collaborating in the interests of some ulterior purpose. For organisms, this is surviving and reproducing; for languages, being means of communication, badges of social identity and markers of solidarity, media for personal and artistic expression. (Lass 1987: xiii)

Little argument is necessary to show that a phrase such as "systems of systems, all collaborating in the interests of some ulterior purpose" invites an analysis in terms of recursive fitness. After all, the focal point is an entity which is defined repetitively at least on one level ("systems of systems") and the notion of "collaboration" for an "ulterior purpose" seems to address fitness. Thus it appears that historical linguistics is a good testing ground for the generality of recursive fitness. We proceed by looking at a debate among historical linguists over the way in which a set of phonological changes should be categorized.

## 1. An example from historical linguistics

Theories about what has happened in the past have a negative taxonomic agenda: among other goals they mean to prevent a misgrouping of the elements of empirical reality or to correct faulty groupings which have already been established. Of course taxonomic theories cannot tinker directly with what has already happened; at most they can demand that lines of inference drawn between extant data points to re-create the past be well formed according to some criteria.[56]

But the aesthetic connotation of a phrase like "well formed" should not be taken to mean that every taxonomic exercise is wholly artistic or academic or in some other way divorced from an objective brand of truth. On the contrary, bad things can happen when taxonomies go awry. In February 1996 at least five California residents ended up in the hospital after eating *Amanita phalloides*, a mushroom species sometimes called "death caps." One of these amateur mycologists arrived at the hospital unconscious and shortly died; another had 40 percent of her liver removed and later underwent a total transplant; the remaining three spent days in serious condition (Goldston 1996: 4; "New Case" 1996: 2). The victims had missed a cue and perceived the death caps as belonging to the class of edible things.

In other words, their taxonomy was faulty, as was Roger Lass's pre-1992 taxonomy of shifts among English long vowels, at least if we believe Lass's own appraisal of his earlier work. For some twenty years he had defended the unity of sound changes which are traditionally subsumed under the title Great Vowel Shift (GVS). During that period his taxonomic assumptions had prompted him to group these changes together, to claim that collectively they fulfilled some criterion which unified them under the GVS rubric. In Lass's view, this unity existing among the individual sound shifts made the GVS itself a single phenomenon. Putting all the changes among long vowel shifts into one basket left Lass's liver intact, but in 1992 he had a change of heart. Motivated by Robert Stockwell and Donka Minkova (1988a), he recanted his belief in the unity of the traditional GVS. Lass remains a devoté of a smaller vowel shift, one which he claims is unified and even deserves to be called "great" (Lass 1992: 153). But the adjective seems an exaggeration, since Lass's new GVS is certainly leaner, even anorexic, compared with the robust, comprehensive traditional version.[57]

This dramatic weight loss highlights another goal of historical reconstruction, one which perhaps describes the historical phonologist's program better than taxonomy: orthopedics. From one vantage point, hard data form a kind of skeleton which is later fleshed out by inference. The taxonomic agenda seeks to get the skeleton right -- to put each empirical bone in its proper place -- before inferences are built around the basic empirical framework. But that is too one-sided a view. Extant hard facts do not belong anywhere in particular prior to the process of theory-building because no skeleton, no taxonomic template, exists until theory has created it. Observational data are moved as the theoretical edifice takes shape, so that the same datum may be used to shore up a rib cage here or strengthen a joint there, depending on what purpose and perspective are dominant at the moment. In other words, where a data point fits and what role it plays there are as much determined by theory as vice versa. Thus the proper orthopedic agenda of phonologists working with the GVS: they should not take for granted that the extant data indicate one shape or another but instead must work with shape and support simultaneously. As phonologists refine and reconsider, there is no telling what shape or even size their theoretical edifices will take on.

It is not surprising, then, that Lass's old (pre-1992) and new GVSs are aesthetically and logically reminiscent of ads for weight-loss pills. In the ads, before-and-after snapshots show a ruddy Rubens-like behemoth nearly popping her seams on one side and a beaming Botticelli's Venus emerging from the same tent-sized pants on the other, or a guy the size of Rodin's Balzac across from one who looks like Donatello's David. It's a whale of a change, but magnitude is insufficient grounds for disbelief. Maybe the pill works as advertised. Maybe the before-and-after models really share the same skeleton, only differently padded. Certainly the early and late Lasses sketched substantially different GVSs on the basis of more or less the same extant hard data points. Lass's early GVS is relatively massive, encompassing shifts among all the long vowel heights in Middle English (ME) and Early Modern English (EModE), while his later GVS is less than half that size. It is not clear that we can choose between such alternative models on the basis of logic alone. Aesthetics may play just as great a role. (A look at movies and magazines makes it clear that my own 1990s American culture would hold up a Botticelli or even a Giacometti as the archetype of human form, but that's no logical reason why a Rubens or Maillol should blush.) However, there is a question about crash diets which we can approach logically: What's in the pill? Lass offers grounds for his epiphany, and on its own terms his reasoning is convincing. But what lies behind his logic?

In this chapter I argue that the unity debate can be characterized in terms of three key metahistorical assumptions, each of which is in turn tied up with a certain concept of fitness. Here fitness is the standard which determines whether a given set of sound changes survives against competing systems. But we will leave fitness out of the picture until near the end of the discussion. The metahistorical assumptions largely determine which systematizations of English long vowel shifts are seen as well formed and which are rejected as improper ways of fleshing out the extant data. Examining the assumptions, for instance, will help us explain Lass's change of perspective. Certainly Lass has a philosophical bent, and over the years he has frequently attended to metalinguistic issues. But I believe the three assumptions to be discussed here have never engaged him fully; like the secret ingredients of the weight-loss pill, they have remained part of the background mystery behind his otherwise well-documented new stance on the unity question. To make sense of Lass's 1992 position we will also review the opinions of Stockwell and Minkova. They, like Lass, do not explicitly discuss the three assumptions below. And as in Lass's case, Stockwell and Minkova's acceptance or rejection of these assumptions is crucial to the positions which they take on the unity problem.

I conclude that Lass need not have recanted, since his earlier belief in the unity of the GVS was consistent with the defensible metalinguistic positions he had held prior to 1992. His new-found devotion to a smaller GVS is likewise defensible on the basis of his present philosophical commitments, but it is distressing to see Lass touting his new view of the GVS as being logically more consistent -- in an absolute sense -- than his earlier position. Better if he had turned art critic and described the aesthetic grounds for preferring his svelte new mini-GVS. This conclusion necessarily calls into question the adequacy of Stockwell and Minkova's 1988 arguments, as well, since Lass names their article as the primary motivation for his change of opinion. In fact any position on the GVS unity debate depends on which metalinguistic assumptions one chooses. Lacking absolute criteria for making such a choice, there is no absolute answer to the unity question.

### (1) First assumption

A phenomenon is unified if its adjacent parts bear an unbroken causal relation to one another. (Negative corollary: If for any pair of "adjacent" sub-phenomena no cause can be found which links one sub-phenomenon to the other, then the overall phenomenon is not unified.)

Discussion of whether the GVS is a single, unified phenomenon raises a semantic question: What do we mean by "unified"? In other words, what criteria must be met before a phenomenon can be said to exemplify "unity"? None of the interlocutors in what I will call the unity debate (or question, problem, etc.) -- the issue of whether the sound changes traditionally subsumed under the title Great Vowel Shift constitute parts of a single whole -- pauses to offer a rigorous definition of this key term. Neither will I, but a few words of clarification may prove useful.

Participants in the unity debate use words such as "unity" and "coherence" and phrases like "*innere Zusammenhang*" loosely and more or less interchangeably.

The same goes for expressions of the opposite quality, e.g., "independence" as opposed to "coherence" (cf. Stockwell and Minkova 1988a: 379). Were we to painstakingly read the relevant literature we might discover all the words and phrases which a given author -- Roger Lass, say -- uses when he wants to assert that a set of sound changes are "coherent" or "unified" or "unitary," or that collectively they have a single "internal structure" (see, e.g., Lass 1976: 52, 53, 58). But such a Roger's thesaurus would only end up looking like *Roget's*, emphasizing that the constellation of synonyms for "unity" is based on the sharing of a common characteristic. That much we could have guessed without doing any reading at all. What matters to us is the substantive question of what that common aspect is.

To save the time and tedium necessary to review every occurrence of the relevant terms, we might propose without demonstration that the commonality we seek resides in two qualities: shared cause and compact time frame. In other words, whether we choose to treat the sound changes traditionally placed under the GVS umbrella as unified rather than independent (or read any of the synonyms mentioned above to characterize this dichotomy) depends on whether we can assign a common cause to those changes and whether the changes happen within a time frame which is sufficiently compact (compact, that is, by standards we will discuss below, under the second assumption). To restate this straw-man proposition more briefly: common causality and temporal compactness make the case for unity, while the absence of either can be taken as reason to deny unity.

### (a) Final versus efficient causality

But the concept of unity remains problematic even if we agree that its foundation is common cause and temporal compactness. That is because the way we describe a candidate for this status -- in other words, the way we draw boundary lines among more primitive, individual phenomena -- is a matter of choice constrained by metahistorical commitments. This is obvious in the case of the temporal compactness criterion: "compact," "short," "brief" and similar terms have no absolute meanings. That is not to say that their definitions are specified "arbitrarily" in the sense of "for no good reason." Rather the reason is arbitrary in an absolute sense because it depends upon a given theoretical apparatus; what one theory takes as a brief time span may be an eternity within another theoretical context. We will revisit this issue below.

The causal criterion may seem more absolute, but it, too, emerges as problematic upon closer examination. What do we mean when we say sound changes are unified because they share a common "cause"? As we saw in Part One of this dissertation, the classic taxonomy of causes is Aristotle's four-part division (*Physics* B.3 194b17 ff.): *material* (What is the mushroom or vowel system in question made of?), *formal* (What essence does the growing mushroom or evolving vowel system express?), *efficient* (Who or what makes the mushroom or vowel system grow in the way it does?), and *final* (To what end state is the mushroom or vowel system progressing?).[58] Each of these types of cause is complicated in itself, but even if we try to accept each at face value, we still have four to choose from. Which kind of cause is relevant to the unity question?

In 1976 Lass appeals to final causation (without mentioning Aristotle) when he attempts to find the motive principle of the traditional GVS:

> ...[W]e must probably adopt something like the notion 'final cause'. The rule-additions are 'effects' prior in time to their 'cause', which is the completed change that we recognize the rules as having contributed to. And at the same time this 'cause' is statable as an antecedently existing condition which determines that the particular rules we find (and no others) should have arisen. I frankly do not see any way out of an ultimately finalistic view of at least some kinds of linguistic change. (Lass 1976: 54)

This appeal to final causation is significant because it means Lass need not be defensive about an accusation we will encounter again below when we turn to Stockwell and Minkova (1988a, b). The allegation is that the unity of a phonetic system such as the GVS is somehow suspect because it is recognized only in hindsight. On the contrary (Lass would have responded in 1976), if the cause of a phenomenon is final, then it is to be expected that the unity of that phenomenon will be perceivable only at the end of the evolutionary process.

We can see the ramifications of this position more clearly if we use graphic terms. In 1976 Lass argued in effect that the right-most column of the chart (1.a) below *caused* (in final fashion) the vowels seen on the left-hand side to change as they did. We should notice that such an argument, one which relies on the concept of final causality, goes beyond positing a push- or drag-chain. Chain arguments rely on the notion that vacancies tend to be filled by something in their vicinity (below them or above them, to continue with the spatial metaphor). In this kind of explanation there may be constraints on how an empty space will come to be filled -- for instance, so that an optimal or minimal phonetic distance is preserved (Stockwell and Minkova 1988a: 358). But in chain arguments there is no way of telling *precisely* what will happen at the terminal extreme of the process. That is because such arguments rely on *efficient* causality: they ask how one state of affairs is causally related to that immediately preceding or following it and not how a series of such proximate changes will eventually end. By contrast, Lass's GVS as final cause determined what the end state of the GVS as process was to be: all the vowels on the left-hand side were, always, becoming those which appear on the right. (Any hint of determinism implied by this conception apparently did not engage Lass in 1976. He also does not address the question of whether we can actually identify the final cause, that is, the true end state of English long vowels or whether the vowels remain in flux even today. Perhaps what he identifies as the final cause is merely a way station along a linear path which is not yet at its end, or perhaps the progress of vowels is cyclical, similar to the big bang-big crunch pattern of some modern Western and ancient Hindu cosmogonies. In 1992, on the other hand, Lass takes care to identify which of his conclusions are "corrigible," but by then he has abandoned talk of final causes. See Lass 1992: 152.)

(1.a)

| ME | | EModE | | | | ModE |
|---|---|---|---|---|---|---|
| i: | → | ei | → | \| | → | aI |
| e: | → | i: | → | \| | → | i: |
| ε | → | e: | → | \| | → | i: |
| a: | → | ε: | → | \| | → | e: → eI |
| u: | → | ou | → | \| | → | aU |
| o: | → | u: | → | \| | → | u: |
| ⊃: | → | o: | → | \| | → | ∂u |

(adapted from Lass 1992: 145[59])

Viewing the right-hand column of this representation of the GVS as a final cause is a choice; it is by no means necessary in the sense of being forced upon us by the objective data or by metaphysical baggage which we *must* carry. To see this, we can contrast Lass's 1976 view with earlier and later appraisals of the GVS's causal status.

Here is a reading which preceded Lass's final-cause theory by some 13 years: "And although we may find the label 'The Great Vowel Shift'...a great help in classifying the series of English vowel changes that have resulted in our modern pronunciations of *stone, house, I, green,* and thousands of other words, we should not be so foolish as to think that the Great Vowel Shift is a *causal* explanation of these pronunciations" (Bloomfield and Newmark, 1963). Certainly we can read Bloomfield and Newmark as cautioning that the GVS describes or systematizes rather than explains in the sense of offering an *efficient* cause for the transformations among long vowels. In the context of their entire book it seems clear that the authors would also deny that the GVS's end-state (assuming we can identify such a terminus) functions as a final cause. In other words, they see no evidence that any identifiable *telos* "pulled" nascent vowel sounds toward it.

Stockwell and Minkova (1988a) are even more explicit in rejecting Lass's use of final causation as an argument for the GVS's unity:

> But is it possible, when one takes the dialect evidence into account, that perhaps the putative unity of the vowel shift is the product of hindsight, and that its "unity" is the linguist's perception, promoted by the simplicity of Jespersen's neat diagram (1909: 231), by the elegance of the Chomsky-Halle rules, and of course by the vested interests of several lifetimes spent trying to establish beyond challenge the internal structure and dependencies within the GVS? (1988a: 356)

There is a telling opposition implicit in this query: on one side, notation, formalism, and Kuhnian social factors militate in favor of unity, but on the other side empirical data -- "dialect evidence" -- will carry the day for Stockwell and Minkova's position, or so they believe.[60] In other words, psychological factors may have led some linguists to impute a unity to the GVS based on patterns which they perceive in past data, but presently available, hard data form the foil which will burst these dreamers' bubble. For Stockwell and Minkova, a unity perceived in hindsight is no real unity at all. That is because they implicitly limit the causal criterion to efficient cause, thereby excluding the final causality which Lass relied upon in 1976.

There is an obvious irony here: Stockwell and Minkova contend that those among their peers and predecessors who perceive the traditional GVS as being unified have in fact been seduced by pretty pictures and other influences not consciously recognized. Yet Stockwell and Minkova seem equally unaware that they reduce the causal criterion to a very specific kind of cause, one among several possibilities. It is clear that they limit their attention to efficient cause, and in the next section we will see that they reject a type of cause even within this category.

This very brief survey of views should suffice to show why the causal criterion of unity is questionable and thus why the concept of unity itself is problematic. Although the task at hand is descriptive -- to analyze what unity means in the context of the GVS debate -- a normative point may help to emphasize where the problem arises: presumably the best we can do is to insist that unity not be denied or confirmed on *unacknowledged* grounds. The Lass of 1976 seems to avoid this sin, since his metalinguistic commitments include conscious acceptance of final causation as a valid unifier. Thus he can place all of the traditional GVS's sound changes under a single causal umbrella and pronounce them unified. But Lass's tormentors, Stockwell and Minkova (1988a, b), and later Lass himself in 1992, limit their attention to efficient cause, without any discussion of alternative conceptions. On this basis they deem the traditional GVS a taxonomic error. Their logic is acceptable, since it is not *necessary* that the causal criterion be fulfilled by a final rather than another kind of cause. But their rhetoric is regrettable in that they purport to reveal a truth whose underpinnings are absolute (i.e., logically necessary) without offering any argument for those metalinguistic commitments.

### (b) Two kinds of efficient cause relevant to the unity debate: inner (participating) and outer (non-participating) causes.

Motivated by Lass's explicit 1976 appeal to final causality in defending the unity of the traditional GVS, we have examined the distinction between Aristotelian final and efficient causes, and we have briefly considered the way final causality functions in Lass's argument. But it will also prove worthwhile to look a bit more closely at efficient causes. In fact we can draw a distinction between two types of efficient cause -- and of course we must then invent a terminology to reflect the difference -- which will help us to better evaluate the positions taken in the unity debate.

The essence of the distinction can be encapsulated in an example. Suppose a juvenile Californian (still too young to have killed herself by eating poisonous mushrooms) carefully arranges several dominoes in a line. She stands each domino on its edge and makes sure that the space between each pair is shorter than the dominoes' height. Call this first state S1. Next the child pushes over the domino on one end of the row and watches the chain reaction with delight: the first domino falls against the second, which then falls against the third, and so on, until not one domino is left standing. The fallen dominoes still form a line reminiscent of S1, but we can indicate their fallen status by naming this second state S2.

Once again the child arranges the dominoes, this time to match the photo of Stonehenge hanging in the living room of her home. (Remember that we're talking about Californians here.) The dominoes constituting the mini-Stonehenge are all standing, but in a pattern different from S1. Call this Stonehenge arrangement S1′. Alas, a sudden earthquake jars the little girl's house violently enough to knock down the dominoes all at once. The roughly circular shape of mini-Stonehenge remains, just as the linear pattern of the fallen dominoes in S2 was still to be seen after their chain reaction, so that there is some degree of continuity despite the obvious change. Call the fallen domino-Stonehenge S2′.

In both transitions, S1 to S2 and S1′ to S2′, we can identify a cause of change as well as various aspects of continuity uniting the before-and-after states (e.g., shape and solidity of form among the individual elements, linear and circular collective patterns). We can also identify different efficient causes associated with each of the two transitions. The chain reaction-type transition from S1 to S2 demonstrates two obvious kinds of cause. First there is the initial impulse, which might be assigned to any number of phenomena -- a fillip of the girl's finger, for instance, or electrical signals crossing synapses in her neural network, or the genetic "hard-wiring" which makes children of our playful species delight in causing things to fall. Initial impulses, in other words, are in the eyes of the beholder; there are lots of ways to name and describe them, and they can be assigned to various "levels" of whomever or whatever we choose to identify as agent (e.g., the girl, her finger, a signal in her neurosystem). The same could be said of the second kind of cause, the more immediate one, though perhaps to a more restricted degree. In state S1 each domino except the last in line is in some sense the cause of the next domino's fall. Whether we call a domino itself or something like the force generated by properties such as its mass and acceleration the cause of its neighbor's fall can remain moot for our present purposes. The fact that most concerns us is that in the transition from S1 to S2, we can identify causes which are *outside* the context of the transition (e.g., the girl's finger) and other causes which are *inside* that context (the force describable in terms of the mass and acceleration of the first domino in the chain). Another way to express this distinction would be to point out that some causes do not *participate* in the transition as described, whereas other forces do. The little Californian *qua* outside cause of the chain reaction does not herself fall over; she is a *non-participating* cause. By contrast, any domino which causes the next in line to fall by falling itself can justly be called a *participating* cause, that is, one which is inside the context of change. So we have two classes of efficient cause: first, "outside" or "non-participating" efficient causes and second, "inside" or "participating efficient" causes.

It should be immediately clear that it is much more difficult to identify inside (participating) causes in the second transition, S1′ - S2′, in which the domino-based model of Stonehenge is destroyed. That is not to say there are no such causes to be found. It all depends on what we take as the locus of change. If the Stonehenge model stands on the floor, and if we identify points on the floor by a system of Cartesian coordinates, then we could say that as the tremors move each such point in three-space with respect to some independent reference mark, each domino's stance is correspondingly disturbed in some way. In that case the vibrating floor participates in the transition. But if we have already specified the context of observation, the locus of change, as including dominoes and nothing else, then the movement of the floor during the earthquake must be seen as an outside (non-participating) rather than an inside (participating) cause.

Let us say, then, that in "chain-reaction transitions" we can readily identify both inside and outside causes, whereas in "earthquake transitions" only outside causes can be easily identified. It is important to remember, however, that the decision to regard a given transition as belonging to one category rather than the other is somewhat arbitrary. We could have chosen to include the floor as part of the context of transition in the Stonehenge example above, in which case the scenario would become a chain-reaction transition and would exemplify obvious inside as well as outside causes.

The domino examples have been a somewhat lengthy digression, but with the observations and distinctions which emerged we can now "unpack" questions about the unity of the GVS with a bit more care. When Lass considers the question of the GVS's unity in 1976, he seems to have been open to earthquake explanations of the kind which André Martinet's 1955 discussion had exemplified.[61] Martinet (as reported by a beguiled Lass in 1976) reasons as follows. In late OE and early ME vocalic quantity began to play an ever-decreasing lexical role. There continued to be various vocalic quantities, but they resulted from structure (e.g., vowels shortened before certain consonant clusters). Martinet names this state "*isochronie*."

A second principle of English vowel organization now comes into play in Martinet's theory: vowels tend to group themselves in lexically significant pairs. Because vowel quantity is de-emphasized in the isochronic state in which the language found itself in the ME period, these lexical pairs were based on quality or nuclear type. However, there was an important exception to this rule: "certain uneven distributions (brought on especially by the quality changes that accompanied open-syllable lengthening) left the two high vowel sets, back and front, once more paired by length" (Lass 1976: 61). Thus the uppermost tier of the long vowels is an exception to the *isochronie* exemplified by the rest of the system. The result is a "natural tendency" to bring the system into symmetry, that is, into the same quality- and nuclear type-based mode of opposing lexically significant vocalic pairs. And so the close vowels began to diphthongize. (Obviously a symmetry principle is assumed to be operative here.)

In our terms, then, the force of change in Martinet's theory comes in the form of a non-participating efficient cause affecting the entire GVS in the same way the dominoes were affected by the earthquake. Martinet's sense of symmetry as a motive force shaping not just sounds but sound systems unites the individual sound changes of the GVS *before* a question of unity is raised. In other words, by choosing a symmetry principle as the motive force of phonetic change, Martinet sets the stage for an earthquake-type account. Here, a symmetry relationship applies to groups of more basic phenomena, not to cases in which each individual phenomenon is viewed in isolation. Thus when a symmetrical group of linguistic phenomena is identified, there is immediately the possibility of an earthquake- as well as a chain reaction-type transition taking place. This is significant because it means that although a conception of final causality sufficed in one of Lass's arguments for the unity of the GVS in 1976, efficient causality fulfills the same purpose in another argument.

Efficient causal explanations can take the context of transition as a unit or as a collection of independent sub-phenomena. Just so, when a neighborhood is leveled by an earthquake, we have warrant to assert that the rubble represents either the entire neighborhood or else the buildings which it once comprised. Either perspective is equally valid, though one or the other might be more useful for a given purpose. Moreover, if someone captured the earthquake's progress on video we may or may not see inside causes for the destruction -- one building falling into another, say -- but we are sure to identify an outside cause, namely, the earthquake itself. In the same way, when the long vowel shifts traditionally subsumed under the GVS are made, conceptually, to constitute a single context at some moment which is specified as being pre-transition, then it is no surprise that the transition itself is seen as a unitary phenomenon. It is worth repeating two phrases of the previous sentence for emphasis: first, elements can, conceptually, be *made to constitute a single context*; second, a moment in history can be *specified as being pre-transition*. The historical phonologist chooses the context of observation. The data may impose constraints, but within those boundaries the linguist has significant latitude to specify not just geographical parameters, but also which phenomena are inside and which are outside the context of observation. Similarly, the phonologist chooses temporal parameters, deciding at what points to "freeze-frame" a sound system, thus yielding before, after and intermediate "snapshots."

How have our interlocutors in the unity debate made these choices? Many of their methods are obvious. Temporal extremes are specified with varying degrees of precision. Perhaps a fairly specific date associated with a particular text is given; at other times phenomena are temporally bounded with less precise phrases such as in Lass's account of Martinet's theory: "in late OE and early ME" (Lass 1976: 61). Similarly, probable pronunciations of the long vowels are offered for points before, after, and during the transition. But linguistic principles can serve to specify context as well. Such principles are sometimes left unspecified. As we saw above, for example, Stockwell and Minkova (1988a) implicitly limit their consideration to *efficient* causality, but they never make this choice explicit.

What kinds of transitions *cannot* be considered unified without analysis of a "chain" relation? Presumably those which fall outside the category of events governed by a final cause and also outside the categories of chain-reaction and earthquake transitions resulting from a single outside efficient cause. In other words, to find unity among sub-phenomena in the absence of a final or an outside efficient cause, we must find an unbroken chain of proximate (inside efficient) causes. But here a problem arises. If in the end we see something analogous to either S2 or S2′ -- that is, a case in which all of the dominoes have fallen over -- there is no way of telling in hindsight whether there was a single outside efficient cause or more than one. For suppose that the dominoes in S1 were arranged such that somewhere along the line the space between one domino and the next is too great for contact to occur. Then the first domino in the line falls over as before, but this time the chain reaction ceases when the long gap is reached. Suppose that the second half falls over as well, only later. Were there two phenomena or one? We cannot say, simply because we cannot rule out the possibility that the same cause (the little Californian, the cat, a thunder clap, a gust through the open window, or what you will) initiated both mini-chain reactions. To complicate matters further, the two mini-chain reactions could be said to share the same outside efficient cause or not, depending on how we choose to describe the cause. If the little girl knocks the lead domino over and watches the first chain reaction, then knocks the first domino of the remaining line over, do we have one efficient cause (a single little girl) or two (two distinct fillips of her finger)? Appealing to time frames will not rescue us here, since arguably no cause acts in null time: we must specify what time interval is sufficiently small such that whatever happens within it is said to be simultaneous with whatever else happens in the same interval.

The phonological significance of the speculations in this section and the preceding one (1.1 and 1.2) can be summarized in the following theses. 1) Causal unity can refer to final as well as efficient causality. 2) The efficient cause of a phenomenon is not wholly determined by observational data. Rather, the efficient cause can be variously described, depending on the theoretical commitments and pragmatic interests of the observer. 3) Efficient causes can be identified as either inner (participating) or outer (non-participating). Although it is easier to identify participating causes in the case of chain-reaction transformations than in the case of earthquake-type ones (in which the outer cause is often more obvious), both types of cause can be described in both kinds of change. 4) Assigning a single cause versus multiple causes is at least sometimes a matter of choice rather than a decision forced by hard data.

### (c) Lass's Zebra and Constellation Fallacies

By 1992 Lass had been sufficiently browbeaten or seduced by the empiricism of linguists such as Stockwell and Minkova that he had taken up their refrain and even embellished it by identifying two fallacies. Lass intends what he calls the "Zebra Fallacy" (1992: 146 ff.) to demonstrate that superficial appearances can lead the unwary observer to overlook a more foundational historical reality. The name of the fallacy was borrowed from an article by Stephen Jay Gould, in which the Harvard paleontologist considers a report that at least one kind of zebra is more closely related to normal (non-striped) horses than to other zebras based on comparison of genetic characters. Here, "kind of zebra" refers to the pattern of stripes; there are three basic stripe patterns and thus three kinds of zebra. The parallel which Lass intends to draw is this: just as the zoological layman mistakenly groups all three kinds of zebra together on the basis of a superficial characteristic (stripe pattern) and in so doing implies that each of the three kinds is more closely related to the other two than to any category of non-striped horses, so the phonological layman or even the careless professional groups all the changes among English long vowels in the ME and EModE periods under the GVS rubric. Why? Because in hindsight the vowel changes form a picture which is most easily grasped if we assume that a genealogical relationship underlies the cosmetic similarity. In both cases, Lass's analogy continues, the layman is proved wrong by a closer, more scientific analysis. Cosmetic similarity does not imply genetic relationship. We will take up Lass's use of Gould in this context in our discussion of the second assumption.

Lass's "Constellation Fallacy" asserts what one would expect from the title. Constellations are not real in the sense of existing in nature independently of an observer. Rather, they are mental constructs imposed on data points. Lass's intention in sketching the constellation metaphor can be read in a weak and a strong version. The weak version would claim that although raw observational data can be systematized, no single scheme is dictated by the data themselves. But in this weak version one *set* of systematizations might be privileged on some grounds, even though no *single* pattern can be chosen over all others. Thus the constellation we call Scorpio might be envisioned as a horse jumping over a house as well as a scorpion, but not as a nativity scene. In other words, a rank ordering is possible, but it must be coarse: some ways of "connecting the dots" are possible, others are not. The strong version would deny this possibility, holding that no system can be preferred to any other, that is, not even a two-tiered rank-ordering is possible.

In 1992 Lass finds empirical grounds to rule out the traditional GVS (particularly the fact that changes at the half-open level and below occurred in a different time frame and seemed not to share an efficient cause with changes among the half-close and close vowels), so it seems clear that he does not intend to defend a position consistent with a strong version of his Constellation Fallacy. (On the question of time frames, see Lass 1992: 153 and Lass 1989; regarding the causal mechanism, see Lass 1992: 150 - 152.) But even the weak version of the Constellation Fallacy -- assuming that is what Lass had in mind -- is problematic. We can imagine a case in which stars (or any other data points) are sufficiently numerous to bear a "real" resemblance to some other object, a resemblance which appears less arbitrary than, say, that between a handful of stars and a scorpion. A counterfeit twenty-dollar bill is not a real twenty, but it seems overstated to claim (consistent with both versions of the Constellation Fallacy) that the similarities evident in the two bills are imposed by the viewer. Of course the very term "resemblance" implies the existence of an active observer. In other words, all resemblances are necessarily observer-dependent if we hold that to resemble presupposes to be seen. But that questionable claim is not at issue here. Nor is the fact that the counterfeit bill was designed to look like the real thing. What matters here is the question of whether one imposed pattern is such a good fit, by some criteria, that all its competitors are ruled out. If Lass answers in the affirmative (as he implicitly does in the rest of his 1992 article), then he rejects his own Constellation Fallacy: what he calls the Luick-Lass version of the GVS is the "real" (though corrigible) one. If he answers in the negative, then he can hardly justify asserting that a single, slimmed-down GVS is "real" after rejecting his original version of the GVS on the basis of hard data in the form of orthoepic testimony and modern dialect evidence. Thus we must read his discussion of this fallacy in a very weak sense indeed. It is merely a caution: "Be careful how you interpret data!"[62]

Since the Constellation Fallacy is so problematic, is it possible simply to throw it out? Lass is very brief in explaining why he points to both the Zebra and

Constellation Fallacies as pitfalls awaiting the unwary historical phonologist. "Why then have these changes [constituting the traditional GVS] been grouped together? The answer lies in another fallacy [the Constellation Fallacy], perhaps less benign than the Zebra because it arises not from naive perception but from deliberate human (creative) activity" (Lass 1992: 147). Perhaps it would be most accurate to say that the two fallacies can be distinguished from one another based on the concept of historicity: there is a historical fact of the matter with respect to the evolution of zebras even though evolutionary biologists and paleontologists are not privy to all the details. By contrast, the only historical aspect in the case of constellations originates in those who, through the ages, have imputed these patterns to various collections of stars. (That the positions of the stars relative to the earth change is not at issue here. These positions vary so slowly that we can take them as being essentially fixed. Even if the changes were rapid, Lass's metaphor would be unaffected.) In other words, the question of whether a group of stars looks like a scorpion or a horse or nothing in particular is synchronic rather than diachronic, whereas the question of how zebras really relate to horses can be approached diachronically as well as synchronically. Presumably the only situation in which we would worry about being caught up in a Constellation Fallacy is one in which the sound shifts comprised by the traditional GVS bear absolutely no genetic relationship to one another. However unlikely such a circumstance may be, we cannot rule it out. If it were true that there is no genetic relationship among any two shifts of long vowels or diphthongs, then it would be fair to say that adherents of the traditional GVS have committed something like the Constellation Fallacy. But we cannot prove that such is the case (just as we cannot prove any negative), whereas there seems to be reason for taking at least some of the shifts as being genetically related to one another.

One possible explanation for Lass's appeal to the Constellation Fallacy is that by 1992 he derives his sense of phenomenal unity wholly from the temporal criterion and from inside efficient causality. This means that he will exclude from the class of unified phenomena any candidate collection of data whose agency seems to be sporadic (see section 2.1 below) or, which amounts to the same thing, whose adjacent individual sub-phenomena cannot all be perceived as causally related to one another. This last qualification means that if A, B, and C are sub-phenomena, then they can only be called part of a unified macro-phenomenon if A bears an inner efficient causal relationship to B, and B to C. Presumably such relationships are perceivable in what could be considered an "instant" compared with the amount of time necessary for the agency of a final or outside efficient cause to become evident. Such an "instant" need not literally approach null time; what is important is that it occur *very quickly* based on someone's sense of what the norm for dramatic linguistic change is. In considering the differences between "U" (upper-class) and "Non-U" speech, for instance, Ross notes that some of the distinguishing features of U versus Non-U are phonetic: "To pronounce words like *ride* as if spelt *raid* is non-U (*raid* was, however, undoubtedly Shakespeare's pronunciation of *ride*). This kind of pronunciation is often called *refained*" (Ross 1962: 99). Ross offers an example of the speed with which such distinctions can disappear or change their significance:

> I am convinced that a thorough historical study of the class indicators discussed above would reveal many present-day U-features as non-U at an earlier period and vice versa. To take an example. In his *Critical pronouncing dictionary and expositor of the English language*, published in 1791, J. Walker is clearly trying to differentiate between U and non-U usage. Yet nearly all the points mentioned by him -- only one hundred and sixty years ago -- are now "dead" and without class significance, in that one of the pronunciations given is today no longer known in any kind of English save dialect. (Ross 1962: 104 - 105)

Ross's "only," qualifying "one hundred and sixty years ago," makes it clear that radical phonetic change can happen very quickly. To the degree that the time span encompassed by a diachronic phenomenon *approaches* this relative "very-quickly" kind of zero, to just that extent the phenomenon itself approaches synchronicity. In such a case it is conceivable that we might fall prey to something like Lass's Constellation Fallacy, in which we take it that a pattern we have created in fact exists.

But if we assume that shifts among English long vowels and diphthongs do bear some genetic relationship to one another (even if not of the direct causal and temporal nature implied by versions of the traditional GVS), it seems clear that speculation about the Great Vowel Shift, if it is to be fallacious at all, must be so in a zebra- rather than constellation-like way. There is a fact of the matter in the questions of whether there was a coherent GVS and, if so, of how it proceeded. That the details must be a matter of speculation does not matter when it comes to recognizing the fact that the shift, if it existed, was not imposed on a collection of synchronically existing data points by later observers, consistent with the Constellation Fallacy. So we need to have a closer look at Lass's Zebra Fallacy. But first let us introduce a further assumption.

(2) Second assumption

Unified structures grow uniformly, i.e. with no long temporal gaps separating adjacent elements of the structure. (Negative corollary: If there is a temporal interruption in the emergence of structure, then what grew before cannot be linked with what developed after.)

To understand how Lass's thought progressed, we need to sample the extremes of his views -- collect before-and-after snapshots of his positions, so to speak. It will also be interesting to speculate as to what motivated him to change his perspective. He cites Stockwell and Minkova (1988a) as the principle catalyst for his change of viewpoint (Lass 1992: 144-145). They assert baldly:

> Changes that are separated by 300 years surely cannot partake of the same "inner coherence". That is, there is ample evidence that Middle English e had become [i:] by 1450, and probably as much as 50 to 100 years (Wyld 1936/53: 207) earlier [a footnote here refers the reader to Prins, Anton A. 1974. *A History of English Phonemes*. Leiden: University Press, 31]; and there is ample evidence that e did not merge with [i:] -- in the dialects where it **did** merge -- until almost 1700. (1988a: 370)

This statement deserves critical scrutiny. If two sound changes, A and B, occur simultaneously, the temporal proximity of their respective occurrences may or may not be due to a common cause. Stockwell and Minkova's description of English long vowel mergers is consistent with more than one sociolinguistic phenomenon -- lower classes attempting to mimic upper-class pronunciations and upper classes adopting new pronunciations to distinguish their speech from that of lower classes, for instance (as in Ross's remark about the fluid relation between U and Non-U above) -- so that it is at least conceivable that the cause of change A differs from that of B even if the two happen simultaneously. Thus temporal proximity is not *sufficient* to demonstrate efficient causal unity. But are Stockwell and Minkova correct in asserting that temporal proximity is *necessary* to phenomenal unity of another kind, and if so, what constitutes "proximity"? If 300 years is too long a period of separation, what about 299 years and 11 months, or, say, 15 years? In citing the 300-year span as a criterion of exclusion, Stockwell and Minkova apparently rely on a faculty of intuition in linguistic matters. In their case such intuition is no doubt well developed. But they owe us a more didactic justification here. As we will see in a moment, another model of historical development in a discipline often cited by linguists finds it common and understandable to observe long periods of stasis punctuated by much shorter moments of dramatic change.

There is a question here which begs to be asked but which Stockwell and Minkova never consider: Could it be that the same social factor or set of factors is responsible for each of the individual shifts? If the answer is affirmative, even tentatively, then we have a common efficient cause -- perhaps internal, perhaps external, depending on one's perspective -- linking individual shifts. Why Stockwell and Minkova are of a different opinion is unclear, again because they did not discuss the possibility of unity based on common social cause in the abstract.[63] They favor a sociolinguistic explanation of vowel change, and that is certainly a defensible position. But Stockwell and Minkova do not treat the possibility that something abstractable as a final or outside efficient cause animates sociolinguistically-motivated changes.

Unfortunately Lass repeats Stockwell and Minkova's unsubstantiated use of temporal separation between sound changes as a justification for denying unity to changes among long vowels. At least Lass's wording is more temperate: "The developments of the lower vowels [i.e., those below the half-close level] however are something quite different: they are not really chain-like, and *they are much later*" (Lass 1992: 152; emphasis added). That statement summarizes Lass's grounds for denying to raisings among these lower vowels the *innere Zusammenhang* which he attributes (in opposition to Stockwell and Minkova 1988a, b) to raising and diphthongization at the top two vowel heights.

Another way of looking at the temporal criterion which plays such an important role in Stockwell and Minkova (1988a, b) and Lass (1992) is to focus not so much on temporal *compactness* (i.e., the size of the time span in which a series of sound changes occurs) and more on what we might call temporal *continuity*: regardless of how long it took for a series of sound changes to play itself out, did the causal agent involved act *continuously* throughout the duration? We will take up this issue in the next two sections.

### (a) Punctuated equilibrium and temporal gaps

It is especially ironic that Lass should have borrowed the title for his 1992 article -- "What, if anything, was the Great Vowel Shift?" -- from American paleontologist and essayist Stephen Jay Gould.[64] Actually, there are multiple ironies here, but to see them will require a minor detour out of historical linguistics proper. The detour is justified by Lass's forays into extra-linguistic regions. He is an eclectic writer in general, but the theory of biological evolution seems especially to engage his attention. As we have seen, he borrows the title of his 1992 paper from Gould; he draws biological and evolutionary analogies in 1987 (e.g. on p. xiii); in 1986 he cites paleontologist George Gaylord Simpson and appeals to geologist Charles Lyell's principle of uniformitarianism (a principle especially beloved of evolutionary biologists).

Evolutionary biology takes it as a given that all organisms which have ever existed are related to each other in some way or another. In other words, there is a historical fact of the matter, even though researchers have no direct access to historical truth but must instead rely on inference. If we assume that all organisms share a single ancestor (as most evolutionary biologists do; q.v. Ridley 1985), then the central question of evolutionary biology is: *How* are organisms related to one another? This question demands a taxonomy, but in fact many taxonomic systems are possible. These systems differ according to their respective principles of grouping, but accounts of evolution are also distinguished from one another on the basis of speed of change and temporal continuity.

Gould is considered a co-founder of the theory of "punctuated equilibrium" and is currently its best-known champion.[65] The theory was developed as a modification of what came to be called Darwinian evolution, which is roughly the notion that species have evolved from a single ancestor through *gradual* evolution by natural selection. Natural selection, in turn, might be defined as the environmental weeding-out of disadvantageous characteristics through the mechanism of genetic inheritance. The key word for our purposes is *gradual*. What is now known as Darwinian evolution and many of its ancestor theories since 1859 were especially anxious to reject saltationism -- intermittent acts of creation-- as well as the Lamarckian concept of evolution by inheritance of acquired characteristics. What such hostile theories shared was the belief that new phenotypes can emerge rapidly, whether by creation or by a non-Darwinian model of evolution. But finding conclusive empirical evidence for a Darwinian model in which change occurs very gradually as organisms adapt to their environments has proved troublesome to generations of researchers. This is especially true of paleontologists, whose primary instrument of empirical demonstration, the fossil record, is more consistent with a model in which staccato instances of change pepper long periods of stasis. Gould sees his theory of punctuated equilibrium as being consistent with Darwinism in so far as he views natural selection as the mechanism driving the punctuated equilibrium model, although some of his colleagues cannot abide any theory whose mechanism acts in such an irregular fashion.[66]

The reason for this digression into evolutionary biology should be clear. Punctuated equilibrium is a hugely influential theory not just because it is consistent with extant empirical data, arguably even more consistent than its gradualist competitor theories. More importantly for our purposes, punctuated equilibrium is influential because it is *internally* consistent (an issue which we will take up in section 3.1 below). For the moment let us leave aside the issue of consistency with empirical data. There is no *a priori* reason why a causal mechanism such as natural selection cannot act intermittently. On the contrary, other aspects of the theoretical apparatus of evolutionary biology warrant the opposite belief. It is theoretically consistent to believe that random genetic mutations *usually* result in phenotypic differences so slight that all such phenotypes are of relatively equal adaptive value within a given environment. Thus when we look at the fossil record (admittedly a chronicle with many gaps) it is not surprising that long periods of stasis are represented nor that dramatic variations seem to have emerged quickly and somewhat discontinuously. Not only is there no surprise here, but such a record does not require us to impute a new causal mechanism to every new instance of stasis or change. One mechanism -- natural selection -- suffices, just as in the case of gradualist versions of Darwinian evolution.

Doubtless we must be careful not to stretch the analogy with phonetic change too far, but we may at least ask whether either of the classical criteria of truth, coherence or consistence, is violated by a theory which suggests that *some* single causal mechanism may explain sound changes separated by 300 (or 3,000) years. If natural selection as a causal principle can bridge immensely longer gaps, albeit in a different empirical domain, there would seem to be no *a priori* reason why the second assumption (A2) must be true.

That brings us to the second irony in Lass's use of Gould's title as a takeoff for his own speculations about the GVS. Gould reported that a single study by Debra K. Bennett had indicated that zebras may in fact not be a "natural kind," that is, a group of organisms more closely related to each other than to any non-zebras. But Gould remained skeptical, concluding that "Bennett's analysis is based upon only three [genetic] characters, none very secure....Unfortunately, we know that at least one of these characters doesn't work well for Bennett's scheme....I conclude that Bennett's proposal [that zebras are not a 'natural kind' because one type is more closely related to non-striped horses than to the other two types of zebra] is interesting, but very much unproven" (Gould 1983: 361 - 362). Elsewhere Gould documents other researchers' painstaking comparison of "folk" and "scientific" taxonomies (Gould 1980: 204 - 213). The results indicate that folk taxonomies, of the same kind which label all striped horses "zebras" and which assume that zebras are a natural kind, in fact are far more likely to coincide with scientific taxonomies than not. Lass and Gould deal with the same basic taxonomic issue -- whether theoretical pictures of reality which do a good job of systematizing observational data are also well-formed or not in the sense of accounting correctly for genetic relationships -- but Gould tends to answer in the affirmative whereas Lass ends up straddling the fence, rejecting the extremes represented by his own earlier position (e.g. 1976) and in Stockwell and Minkova (1988a). The temporal gap issue is just one of Lass's motivations for rejecting what amounts to a folk taxonomy, that is, an assertion of relationship which he thinks is based on the superficial unity of, say, Jespersen's 1909 diagram of the GVS.

A final problem with using a temporal gap to disqualify certain shifts within the GVS is epistemological. It may be that gaps of some magnitude are necessary to diachronic study. Lyons puts it this way:

> Moreover the notion of diachronic development between successive states of a language makes sense only if it is applied with respect to language-states that are relatively far removed from one another in time...If we take two diachronically determined states of a language that are not widely separated in time we are likely to find that most of the differences between them are also present as synchronic variation at both the earlier and the later time. From the microscopic point of view -- as distinct from the macroscopic point of view that one normally adopts in historical

> linguistics -- it is impossible to draw a sharp distinction between diachronic change and synchronic variation. (Lyons 1981: 58)

Stockwell and Minkova's appeal to dialect-borrowing as the mechanism of change among long vowels (1988a: 379; 1988b: 416) is particularly interesting in the light of Lyons' remarks. Stockwell and Minkova focus on various individual dialects in the South, whereas Lass concentrates on a narrower sampling of Southern speech. (Again, the phonologist chooses the context of observation, and the choice affects the answer to the unity question.) "Our view is that phonetic variation within the domain of a phonemic category **always** exists....The reality is this: variation is rampant in all languages at all times. The interesting question is not where do variant forms come from, but why do some of them, and not others, get singled out and valued within subcultures. This is the same as the question, Why do some fashions become fashionable? We don't know" (Stockwell and Minkova 1988b: 416; their boldface). Thus Stockwell and Minkova push the unity debate into the realm of sociolinguistics, whereas the Lass of 1976 interests himself in non-social factors driving sound change. But what constitutes the proper scope of analysis for each of the two approaches? Lyons' "microscopic" and "macroscopic" are clearly not rigorously defined, perhaps even indefinable in a way which could apply to all tasks of historical phonology. What makes this problematic is that when we approach a phenomenon such as the 300-year gap separating two "halves" of the GVS, we might choose to call the context macroscopic, meaning that we really are looking at two distinct phenomena. On the other hand, an explanation such as Lass's 1976 final cause or Martinet's 1955 *isochronie*-plus-symmetry principle can effectively take the gap as microscopic -- as a period in which variations among dialects are sorting themselves out.

It seems doubtful that in 1992 Lass believed the temporal gap between changes at the upper versus the lower long vowel heights was *sufficient* to make the traditional GVS disunified. The gap was simply consistent with the efficient causal discontinuity between upper and lower vowel heights. Stockwell and Minkova's strong wording in 1988a indicates that they placed rather more weight on the temporal argument. Whatever their positions with respect to sufficiency in this case, Lyons' remark shows why employing temporal criteria in discussions of diachronic unity is problematic. Let us suppose that data would lead us to divide the GVS into just two regions, say North and South. (This is coarse, but for the purposes of argument we can grant the hypothetical for a moment.) Further let us say that a GVS-like change occurred in both regions, but much earlier in the one than the other. We might propose, for instance, that a GVS-like phenomenon occurred in the North long before anything similar happened in the South. Now the question is, must we therefore hold that nothing unites the two shifts?

To answer the question, it may prove worth our time to try to see what Stockwell and Minkova (1988a) and Lass (1992) are saying in abstract terms about the temporal criterion of unity. To this end we can do a thought experiment. Consider a diagram much like that developed by Jespersen in 1909 or (1.a) on p. 5 above, only leave the particulars out. That is, forget for the moment which sounds are changing and concentrate only on the fact that some series of displacements is occurring. (We could of course neglect to stipulate that the changes affect sounds; instead, this thought experiment could simply specify that some group of phenomena displace one another, and that the phenomena are homogeneous in some aspect. All might be visual phenomena, or bodies moving in a two- or three-dimensional space, etc. But there is no need for that level of abstraction here.) The abstract diagram, then, could look like this:

(2.a) $$[x_1] \rightarrow [x_2] \rightarrow \dots \rightarrow [x_{n-1}] \rightarrow [x_n]$$

Now suppose we go further and label each of the arrows to reflect the time period in which the change took place (assuming that the process has come to an end). We could picture the situation in this way:

(2.b) $$[x_1] \xrightarrow{t_i} [x_2] \xrightarrow{t_{i+1}} \dots \xrightarrow{t_{i+k-1}} [x_{n-1}] \xrightarrow{t_{i+k}} [x_n]$$

And now the crux of this thought experiment: What criteria would lead us to regard the phenomenon pictured in diagram (2.a) as a single phenomenon? What about (2.b)? It is not clear that an unequivocal answer is possible. However, it seems certain that the case for phenomenological unity is *strengthened* if we can find a causal connection. But as we saw above there is more than one way to conceive of causal relationships. To say that [x1] is the *efficient* cause of [x2] implies that [x2] does not exist unless [x1] either exists or has existed. The same would be true throughout the entire chain. It also helps if there is a clear-cut sequentiality:

(2.c) $$\text{For all } i,\ t_{i+1} > t_i$$

If condition (2.c) is not met, a quite subtle causal relation would have to be posited before one could conclude that the situation in (2.b) is a single phenomenon. In other words, if the phenomena depicted happen "out of order," then the challenge of finding a unifying causal explanation becomes more difficult. Here "order" means something fairly obvious but something which depends on the type of phenomena under consideration. For instance, in long vowel shifts occurring between ME and EModE, there is a diagrammatic order imposed by the sharing of sounds in the "chain" of shifts. One could write

(2.d)
ME [ɛ] → EModE [e:]
ME [e:] → EModE [i:]
ME [i:] → EModE [ei]

(cf. Lass 1992: 145). Alternatively, one could describe the state of affairs this way:

(2.e) ɛ → e: → i: → ei.

It is not clear that the transition from (2.d) to (2.e), which amounts to exercising what an algebraist would call the property of transitivity, is natural rather than diagrammatic. Nor is it beyond doubt that condition (2.c) must be true before such a transition can be made. To repeat, however, it seems axiomatic that it is easier to go from (2.d) to (2.e), and then to understand (2.e) as a single phenomenon, if condition (2.c) is met.

Sequentiality is important here, the timing of individual shifts is less so. That is, one can speak of certain phenomena -- geological ones, for instance, or developments in the realm of evolutionary biology -- as being driven by the same motive forces for eons, even though long periods of stasis occur. What would disturb geologists and evolutionary biologists is a jostled organization. In the case of long vowel shifts, sequentiality is arguably present, thus making arguments based on final and outer efficient cause possible without lengthy treatment of the "inner" arrangement of vowels. Stockwell and Minkova along with the Lass of 1992 need to offer us more justification when they claim that any "long" temporal gap precludes unity.

### (b) Stockwell and Minkova (1988a)

To help us make sense of the second assumption a further digression is in order. We have already seen some of the key contentions made by Stockwell and Minkova (1988a, b). They begin by asserting that there is no scholarly consensus on several questions raised by Luick with respect to what they call the "English Vowel Shift." It is worth noting that this phrase itself is Stockwell and Minkova's first reactionary salvo in the battle against the notion that the long vowel shifts constituting what we normally call the "Great Vowel Shift" (GVS) in fact possessed enough unity to warrant that august title. Thus to name high vowel raisings and diphthongizations with the traditional "Great Vowel Shift" would be to beg the question in favor of the position which Stockwell and Minkova intend to oppose. Apart from that rather oblique attack through nomenclature, the article's brief introduction (in which five questions relevant to the GVS are delineated) may well mislead the careless reader into believing that Stockwell and Minkova intend merely to discuss five major aspects of this English Vowel Shift.

It would be more correct to say that Stockwell and Minkova aim to fight on five fronts (corresponding to the five problems) against the orthodoxy which claims an internal coherence for the traditional GVS. A close reading of the article makes this obvious. For instance, in a footnote to their discussion of the first problem, the authors are explicit: "Since our purpose here is to call into question the existence of this 'chain of events' [i.e., the traditional GVS] as a unified event, the issue of what caused the diphthongization of high vowels in the first instance, unless, of course, it already existed in Old English (as we believe), remains somewhat peripheral to our argument" (Stockwell and Minkova 1988a: 381 - 382).

Thus the first four discussions are really skirmishes meant to prepare for final victory in the fifth and final question, which deals explicitly with the unity or independence of changes among English long vowels. Stockwell and Minkova name these battle fronts (1) "the inception problem" (Lass's 1976 term): Can we reasonably talk about the Shift's impetus? If so, what was that initial motivation? (2) "the merger problem": How do we evaluate Luick's metaphor of vowels as particles in the mouth and the follow-on theory of "psychological distance between phonemic entities" (Stockwell and Minkova 1988a: 355)? (3) "the order problem" (again a phrase coined by Lass in 1976): Did the shift occur in stages, and if so, how can we delimit them?

(4) "the dialect problem": What role did in-gliding and out-gliding in dialects outside London play in the shift (if any)? and (5) "the structural coherence problem": Does the "English Vowel Shift" have an internal structural unity? A short sketch of Stockwell and Minkova's treatment of the first two of these issues will help show how they stack the deck before tackling the final question. We will consider Stockwell and Minkova's handling of the order and dialect problems below.

Stockwell and Minkova's explicit consideration of the inception problem is brief -- scarcely more than a page long -- and for our present purpose their approach and conclusion may be summed up even more briefly. The authors hold that this is a "pseudo-problem as posed in all earlier work as well as in recent proposals" (357), and they promise to provide an argument for this position later in the paper. For the present they have simply observed that there is a tradition of explanation based on prosodic principles. They place Luick in opposition to this tradition and remark that "he did not attempt to deal with it" (i.e. with the inception problem) and that his approach of focusing on vowel raising as prior to diphthongization somehow "deflected the attention of anglicists away from the inception issue" (ibid.).

As for the merger problem, Stockwell and Minkova offer us Luick's 1932 speculation that phonemes, viewed as the units "which play the decisive role in the language and its development,"[67] are separated by a sort of psychological space so that each phoneme has its territorial imperative. This psychological "space" cannot be trespassed by other phonemes without the risk that the invaded phoneme must move (Stockwell and Minkova 1988a: 357). Thus Luick is presented as defending two major assertions: first, that phonemes are spread evenly across a phonetic space, and second, that they move in an interdependent manner such that the interstices are preserved.

Luick gets no argument from Stockwell and Minkova on the first count. On the contrary, they assert that modern research bears out a picture of "psychologically real and approximately equal spacing of vowel phonemes within [available phonetic space]" (358). But with regard to the second contention -- that the interstices tend to be preserved -- Stockwell and Minkova are less convinced. They find evidence that mergers are rather common. (This does not contradict Luick's first claim, since even spacing is observed among existing vowels, that is, among those which have not merged.) At best we could say that Stockwell and Minkova find ambiguous evidence for what they take to be Luick's claim that there is a psychological tendency to preserve phonemic interval, or put another way, to avoid mergers. The primary problem is that vocalic contrasts decrease in some languages if we consider [LENGTH] and [GLIDING]. Omitting these features would make Luick's theory "generally correct, though even within the area of the vowel shift as traditionally conceived the merger of *e* and *e* at [i:] (*see, sea*) is one counterexample, and the merger of *a* with [ei] (*pane, pain*) is another" (Stockwell and Minkova 1988a: 359).

So much for the Luick of 1932. Stockwell and Minkova now visit the earlier (1898-vintage) Luick, whom they describe as supporting "more a particle-based theory than a feature-based theory" (ibid.). In this period Luick considered dialectal evidence demonstrating that wherever *u* failed to diphthongize, *o* fronted rather than raised. According to Stockwell and Minkova, Luick's conclusion that raising *o* pushed *u* to diphthongize is the first example of a "conspiracy" theory in which one sound change is offered as the cause of another, ostensibly independent change. (Luick's observation was heartily endorsed by Lass in 1992.)

But what's wrong with such a conspiracy theory, assuming Luick's evidence is sound? Here is Stockwell and Minkova's criticism in a nutshell:

> "But it remains true that Luick did not demonstrate that diachronic preservation of vocalic contrasts is a necessary consequence of perceptually equal synchronic spacing of vocalic units. No one else has demonstrated it either. Lacking such a demonstration, it continues to provide a weak foundation for a theory about change within vowel systems" (Stockwell and Minkova 1988a: 359).

Here Stockwell and Minkova seem to have missed the significance of Luick's argument *with respect to long vowel raising and diphthongization.* It would be merely convenient if we could predict the diphthongization of /u:/ following the raising of /o:/ *a fortiori* from a general and proven preservation principle. But whether such a general principle has been proved or not, Luick's evidence at least points to a *local* preservation principle: in this very specific case, there was preservation of vocalic features rather than merging. That suffices for Luick and later Lass to make their arguments for unity in the limited context of the GVS.

Stockwell and Minkova consider in some depth the criticism of Claude Boisson, who in 1982 asserted that dialectal exceptions vitiate Luick's claim that raising of the half-close vowels pushed the ME close vowels to diphthongize.[68] In brief, the authors conclude that Boisson's bases his first attack -- the claim that the high front vowel sometimes diphthongizes without the half-close front vowel having raised -- on two informants of dubious value whose survey responses appear in the *Survey of English Dialects* (SED). One of these informants (39.3) has [ai] where we would expect [e] or [i] (e.g. in *yeast, reel, niece, agree*). Stockwell and Minkova's verdict: "for these to have [ai], they should have been raised **very** early and then diphthongized with the original i, and if that happened the whole point of citing this dialect area is vitiated" (Stockwell and Minkova 1988a: 361; their boldface). The other informant (39.1) loses credibility because his responses diverge markedly and inconsistently from dialects in the near vicinity. Boisson's second argument was that in Gloucestershire we sometimes see merger rather than displacement in the top two back vowels. Stockwell and Minkova reject this interpretation of the data. They allow that ME u "is not strongly diphthongized," but they assert that "there is **always** a contrast between the surviving reflexes of u and the surviving reflexes of o" (Stockwell and Minkova 1988a: 363; their boldface). In other words, there is no merger and thus there is no strong counterexample to Luick.

Stockwell and Minkova leave us hanging. Whether or not Luick was correct in asserting that a psychologically-defined phonetic space tends to exist (or in other words that mergers tend not to happen), there are no unequivocal dialectal data to the contrary. The dialects show (with the caveat noted above) that the ME half-close vowels raised and the ME close vowels diphthongized. If Luick is to be contradicted on any point, then, it is on the temporal and causal questions: Did raising precede diphthongization? Did raising cause diphthongization?

The moral of the story will become the refrain of our discussion of the third assumption. Unity is in the eye of the beholder: it is dependent on whatever theoretical baggage we choose to carry. If our interest is in raising *per se*, then Stockwell and Minkova may well be right when they claim that long vowel changes among English dialects in general fail to exemplify raising. Similarly, if our interest is in finding a type of change which was unique to ME and EModE vowels, then the GVS may not fit the bill, depending on how we read the dialect evidence. But those are very specialized claims about particular criteria of unity. There are other such criteria. Even Stockwell and Minkova find enough consistency among the changes which they believe did occur to recognize and articulate principles of change (q.v. section 3.2 below). If we equate objective independence among the sound changes with total randomness of each sound change with respect to all others, then the long vowel shifts are not independent. That they may not be unique is a different matter.

There are a number of ways to distill Stockwell and Minkova's myriad claims, and each method yields a different thesis. Here are two obvious possibilities. I take these two theses to be the extremes; other interpretations may yield different bottom lines, but these will fall somewhere in between the two boundary theses, T1 and T2.

T1: We cannot, with absolute certainty, claim that any of the criteria of unity discussed above (e.g., final or efficient causal unity) was operative in any of the shifts commonly referred to as the GVS.

If this is what Stockwell and Minkova mean, then they are offering us a very uninteresting disunity, couched in a negative form which rightly left Lass (1988) nonplussed. Of course we cannot offer a mathematical demonstration that a single, organic causal principle of raising and diphthongization operated effectively and without interruption to move Middle English vowels to their present positions. (The continuation or cessation of such a principle presents further problems. If the principle has ceased to exist, meaning that vowels are stable -- something which seems to be untrue -- then we have the burden of explaining why it no longer operates. If it continues to be operative, then it is difficult to name the principle in terms of its effects -- e.g. raising -- since we don't know what its ultimate endpoint, assuming it has one, will be.)

The standard of proof in T1 is impossibly stringent for a historical science. T1 uses the phrase "absolute certainty," and it may be objected that Stockwell and Minkova are by no stretch of the imagination *that* strict. But their rhetoric can conscientiously be read as demanding an extremely rigorous standard of proof. It is essentially the standard used in the United States for criminal murder cases: the evidence has to point to one and only one conclusion *beyond a reasonable doubt.* If one can find any grounds for such a doubt -- another theory compatible with all the data, say -- then the accused must not be convicted and the sound changes may not be confined under a rubric such as the Great Vowel Shift.

There is another possible reading of Stockwell and Minkova's thesis:

T2: None of the principles of unity discussed in this paper (causal, temporal, systematizing) is evident among any two of the individual sound changes traditionally subsumed under the title GVS.

Whereas T1 was worded with such circumspection that there is no reasonable way to meet its standard of proof, T2 is a recklessly strong -- and obviously false -- thesis. Contrary to T2, it appears that a push chain may indeed be operative at the ME close and half-close levels, as Luick (1898) and Lass (1992) suggest. Whether any dialects contradict this general observation is unclear. Boisson (1982) as reported by Stockwell and Minkova (1988a) thinks such contradictory evidence exists; Stockwell and Minkova disagree. Even if Boisson were correct, one could interpret his counterexample as an exception to a rule which does indeed indicate a kind of causal unity.

We have seen Lass's 1976 concept of the GVS as a final cause, in which vowel sounds are drawn toward a pre-existing final state. In such a scheme the vowel system is self-contained and self-motivated. There is no obvious latitude for choice on the parts of those who use the vowel sounds in question as part of their language. In stark opposition to this understanding, key words in Stockwell and Minkova's description of the GVS are "conscious" and "arbitrary," as in their quotation of Dobson (1955): "'[T]he elimination and avoidance of others [i.e., other pronunciations] must inevitably, and quite properly, have been a conscious and indeed often arbitrary process'" (Stockwell and Minkova 1988a: 378-9).

Another key contention of Stockwell and Minkova is that phonetic space arguments are based on an ambiguous foundation; dialect borrowing is preferable. Stockwell and Minkova take special offense at arguments which rely on vague assertions about "preservation of phonetic space," especially those linked with "perceptually non-salient variation along with imperceptible directional drift" (cf. Stockwell and Minkova 1988b: 415). The authors prefer concrete justifications which they can reach out and touch, so to speak, which means they must stick with empirical phenomena the observation of which requires little or no inference -- in other words, hard data from dialects. For that reason Stockwell and Minkova view their position as "realism ... versus abstractionism or idealization" (1988a: 379).

In response to Stockwell and Minkova's disparagement of theories of intradialect sound change -- recall that they favor dialect-borrowing as the mechanism of vowel change -- Lass cites modern research (including Labov) which he takes as indicating that intradialect changes do indeed take place and that Hockett's theory of phonetic space accommodates the observed data. Moreover, Lass believes that Stockwell and Minkova's dialect-borrowing theory begs the question because it fails to account for why there are phonetic differences among dialects in the first place. What we have, then, is a number of disputes over "meta-commitments" rather than over hard data.

### (c) Martinet and temporal continuity

There seems to have been a time in Lass's career when he would have rejected the second assumption out of hand. In 1976, as we have seen, he was fascinated with what he called the "essentially Praguian" account of the GVS's inception developed by André Martinet. Recall Martinet's contention from section 1.2 above: in late OE or early ME, long vocalic quantity tended to lose its lexical significance, thereby making quality (open/close) and nuclear type (vowel/diphthong) more important, lexically speaking. The tendency toward what Martinet termed *isochronie* had created opposing long vowel pairs based on quality everywhere but among the highest vowels (i and u). This asymmetry caused a shake-up of the system as a whole, resulting in diphthongization at the high level and raisings everywhere else. In other words, a symmetry principle was the mechanism of change, the causal force, and the object on which the mechanism acted was the system of early ME long vowels as a whole. This view stands in sharp contrast to the perspective evident in assumption A1 above, where an individual part of a system -- a single sound change, for instance -- must be the cause of change in another individual part before the two parts can be called unified. Martinet's perspective likewise contradicts A2, since his theory does not demand an unbroken chain reaction, that is, temporal continuity among inside efficient causes. (Temporal continuity also loses its importance when a final cause is posited, à la Lass (1976), since the final cause is held to be operative at all times, regardless of how sporadic the changes which it governs may appear to be.)

Lass remained unconvinced that Martinet had gotten all the details corrects. But Lass liked the *type* of explanation which Martinet offered. Here is how he characterized Martinet's basic strategy:

> The GVS is to be explained in terms of processes which affect, in some way, the system AS A WHOLE: it is not a mere succession of local events whose unity is the product of hindsight. The theory itself gives us *a priori* terms for explanation. So that every individual change that was part of the process was brought about, not only as a 'consequence' (in some sense) of a preceding change, but specifically as a consequence of that change having disturbed a pre-existing state of well-formedness, defined on the entire vowel system as a structure (Lass 1976: 62; his capitalization).

In 1992, when Lass denies the unity of the GVS which he once professed, he does not even consider such "Praguian" arguments based on symmetry principles. The question is, why not? Martinet's contention that late OE or early ME tended toward *isochronie* seems defensible. For instance, Platt and his co-authors summarize the development of vowels in so-called New Englishes, which include "Indian English, Philippine English, Singapore English and African Englishes of nations such as Nigeria and Ghana," this way:

> ...[W]e can see that there are some general tendencies which are shared by some or all of the New Englishes:
> (1) a tendency to shorten vowel sounds;
> (2) a lack of distinction between long and short vowels;
> (3) a tendency to replace central vowels by either front or back vowels;
> (4) a tendency to shorten diphthongs and to leave out the second sound element in a diphthong (Platt *et al* 1984: 3; 37)

Tendencies (1), (2) and perhaps (4) seem quite consistent with the general trend toward *isochronie* which Martinet perceived: vowel length tends to be less significant than quality. So at the very least his theory seems to be consistent with a Lyellian principle of uniformitarianism.

It is obvious that a theory of vowel change relying on final cause can accommodate temporal discontinuity. Martinet's theory demonstrates that an explanation relying on efficient cause (in this case, an outside efficient cause) can succeed in the face of temporal discontinuity as well. We can conclude that the Lass of 1992 and his catalysts, Stockwell and Minkova (1988a, b), rely on a "meta-commitment" to temporal continuity which they have not explicitly justified and perhaps not even recognized.

### (3) Third Assumption

The third and final assumption goes like this: In general, unity or disunity is a quality of real phenomena. In particular, unity or its opposite is a quality of the physical sound changes traditionally called The Great Vowel Shift. This means that the GVS unity question is an empirical one. (Negative corollary: Questions of unity do not apply to bodies of knowledge consisting of inference as well as data derived from "objective" observation. Rather, unity or disunity inheres in empirical fact. This means that unity can emerge only where a *correspondence* condition of truth applies; unity cannot be asserted on the basis of intra-theoretical *coherence*.)

As in the case of the first assumption, the wording here raises semantic questions. "Unity" was the key term in the first assumption. Here we must ask ourselves what "phenomenon" and "real" mean. It will be helpful if we first attend to a prerequisite issue:

### (a) Correspondence vs. coherence understandings of truth

When we attempt to construct a good taxonomy, whether of shifts among English long vowels or of mushrooms, we are trying to reflect truths about nature. To ask whether that mushroom over by the rock is poisonous or edible is to be interested in a present fact, whereas to question whether *mate* belonged to the class of words with [e:] as long vowel in a given dialect in 1600 is historical. But in either case we are after the truth of the matter. We can restate any such taxonomic question by first formulating a proposition and then prefacing it with the phrase "Is it true that...": Is it true that all the mushrooms in my basket are of the same type? Is it true that changes among all ME long vowels share the same initial cause? But before we can answer these questions we must have criteria by which to judge truth and falsity. Presumably any historical discipline, diachronic linguistics included, must acknowledge a *correspondence* sense of truth: there must be a way of appreciating propositions which accurately express observations of "outside" data, that is, data which cannot be considered the creations of theory-based inference. But data are often sparse in historical disciplines. The gaps must be suffered or else filled in by guesswork. As more is inferred from less, the theoretical edifice grows in proportion to the observational foundation. It then becomes important to analyze the theory itself for possible inconsistencies. To the extent that few are found, the picture painted by theory-plus-data is coherent, and thus there is a *coherence* sense of truth in addition to the correspondence sense. Although this is a coarse and problematic sketch of two long-recognized truth criteria, it will suffice to suggest why we should expect historical linguists to demonstrate extreme respect for theory and its fruits and therefore for a coherence sense of truth: without coherent theory there is no history.

Respect for theory is exactly what we find embodied in the early Lass and Stockwell. In 1969 Lass edited an anthology of articles in which Stockwell plays a prominent role. In his preface, Lass approvingly quotes from one of Stockwell's articles in the volume:

> History is not an account of facts but of relations that are inferred to have existed between supposed facts. It is not at all easy to make a crucial observation as to what a particular fact is, or to discriminate between facts and inferences. The 'facts' of historical scholarship are often simply useful hypotheses that in turn relate, by rough rules of inference, a variety of secondary 'facts' to each other. The most insightful accounts of historical events often turn out to be intricate webs of suppositions and inferences removed at many steps from the citable data on which the conclusions ultimately rest (Lass 1969: 3; Stockwell 1969: 228).

Lass, too, implicitly defended a coherence theory of truth which allowed that the value of an historical statement may turn out to be proportional not to its "accuracy" in reporting empirical facts (which often are not directly recoverable) but to its consistency and explanatory power (Lass 1969: 3).

It seems likely that in 1969 Lass and Stockwell would have rejected all three of the assumptions discussed in this chapter. Because they recognized the central role played by inference, they might also have had to acknowledge the possibility that a theory asserting the unity of the GVS *could* be coherent. On the other hand, both Lass and Stockwell imply that empirical data (i.e., directly observed as opposed to inferred data) come too far and few between for us to prove a correspondence between the proposition "The GVS is a single phenomenon" and the extant empirical data. Thus, given Lass's and Stockwell's metalinguistic commitments c. 1969, unity emerges from coherent theory rather than some correspondence resemblance between phonetic knowledge and extant data. Of course data cannot be left entirely out of the picture. Part of the job of building coherent theory involves framing statements about unity which *correspond* to the extant empirical data, but such statements by themselves would never suffice to create a full-blown theory. On this view events can possess a unity based on final cause, since a theory relying on final explanation certainly can be coherent. Alternatively their unity can be based on efficient cause when they share a common initial impulse or if they move each other in a chain reaction. But in either case the unity is a product of theory as well as of observation, and the unity is described with propositions whose power is derived not primarily from their correspondence with an extra-theoretical reality, but rather by their coherence with one another.

All of this is not to say that any linguist who recognizes a coherence sense of truth is unfettered by empirical data. In fact no one has *carte blanche* to declare collections of data unified. Existing data always constrain the range of possible theories, since coherence is expected not just intratheoretically but also between empirical data and theory. And that last sort of coherence -- between data and theory -- amounts to correspondence. But what Lass and Stockwell tell us in 1969 is that the creative latitude of theorizing is much greater than many thinkers of the previous generation, particularly the logical positivists, believed (cf. section 3.5 below). This latitude necessarily results from the paucity of data, but it is not a willful, arbitrary neglect of facts gleaned through observation. It is just that sparse data can be consistent with numerous divergent theses. Or in the language of contemporary philosophy of science: data *underdetermine* theory (q.v. section 3.4). (The question of whether this is always the case lies outside the scope of this chapter, but in passing we will consider Kuhn's and Quine's opinions below.)

Neither Stockwell nor Lass explicitly mentions correspondence and coherence criteria of truth, but the distinction is useful for understanding their 1969 position. What we need next is a way of understanding why two phonologists who seem to share the same philosophical basis come to diverge so sharply from one another in later years. Their parting of ways stems from something more than differing interpretations of data; rather, their apparent philosophical harmony masks a fundamental disagreement as to what phonetic structures are at our disposal. Lass leans toward legitimizing pragmatic conceptions from whatever source, while Stockwell grows increasingly suspicious of any entities which cannot be directly perceived. This difference in basic philosophy is evident, for instance, in Lass's 1976 recognition of sound systems as "primitives" in phonetic theory, a decision which the Stockwell of 1988 apparently rejects.

### (b) "Phenomenon" as that which is unified

We can try to employ our earlier observations on the meaning of "unity" as a shortcut to resolving the meaning of "phenomenon." Further, we can consider the ability of certain schemes to systematize data well or poorly. Heuristically, "phenomenon" might be used to indicate any of the following: something that we can describe systematically; something which has a single cause, e.g., objects falling down (i.e., towards a center of gravity); something that occurs in a compact spatio-temporal context; something which occurs uniformly and predictably. (Predictability, as Lass points out in his 1986 paper, can be attributed to phenomena which have already occurred as well as to those which have not.) Presumably uniformity and predictability can be characterized in terms of common causal and temporal qualities. In other words, we might take "phenomenon" to mean a *unified* thing, relying on the discussion above to determine what we mean by unity. If we take this course, however, the problematics of unity discussed above inevitably affect the definition of phenomenon.

That leaves us with phenomenon as a sort of template -- a means of systematizing data which in fact are not genetically related to each other. We have seen above that Bloomfield and Newmark consider the GVS in this light. Responding to Stockwell and Minkova (1988a) Lass, too, suggests that the GVS has a unity as systematizer independent of any causal or temporal unity it might possess:

> As a way of defining both 'Early Modern' as opposed to 'Middle' vowel systems in English, and as a way of separating the 'True North' from the North Midlands and the rest of England, the GVS is still at very least a useful conception and does -- however you feel about the metaphysics of its *enchaînement* -- represent a powerful geometrization of a major quality-reorganization. (Lass 1988: 408)

Further, Lass makes a comparative appeal which seems closely tied-up with a vision of the GVS as effective template for systematizing sound changes: "The only real substantive claim involved in the classical GVS 'ideology' (as I'm quite content to call it) is that -- whatever else has been and still is going on in the history of English vowels -- there was one particular set of late mediaeval shiftings that was more coherent and more potent in effect on the system as a whole than others" (Lass 1988: 407).

But even the understanding of "phenomenon" as a systematizing rather than a causal explanation is problematic when applied to the GVS. This is especially evident in Stockwell and Minkova's (1988a) treatment of what they call the "dialect problem." Their contention is that believers in the unity -- the "phenomenon-hood," so to speak -- of the traditional GVS have painted a false picture of the unidirectionality of vowel changes. They claim erroneous representation resulted when linguists took into account only a limited subset of the relevant dialects. A properly comprehensive survey of the dialect information yields a much different picture, Stockwell and Minkova continue, one in which vowels change in several directions. Thus there is no general raising of long vowels, nor can we find consistent movement in any other direction (1988a: 371 - 375).

This is not to say that there is no uniformity at all among the sound changes traditionally made part of the GVS. On the contrary, Stockwell and Minkova cite general rules governing these changes. We see this in their treatments of what they call the order and dialect problems. Two arguments are to be found here, a positive and a negative one. The negative argument is that there is no coherent way to specify the order of the sound shifts traditionally subsumed under the GVS because there was no unidirectionality among the shifts themselves. Instead -- and now we come to the positive argument -- the shifts mirror a set of principles which are instantiated through the process of dialect-borrowing and which led to multidirectional shifts. These principles govern sound changes among all Germanic languages and may have done so even in Proto-Germanic (Stockwell and Minkova 1988a: 371). The principles are:

> (1) Alternation between V: and VG, usually at the vowel extremes (e.g., between [i:] and [Ii], or between [i:] and [Ie]. (2) Tendency toward expansion of the distance between endpoints of diphthongs (i.e., dissimilation) to achieve perceptual optima....(3) Remonophthongization, or at least movement in that direction, after reaching diphthongal optima ... (e.g., [æu] > [æo] > [æ∂] > [æ:] *house* in the American Tidewater South...) (4) Vowel alternations in adjacent dialects (certainly historically, usually also geographically) are as the rook moves: i.e., [æu] may alternate with [ɛu] or [au], but not directly with [Λu] or [⊃u]. Not uncommonly, **for social but not linguistic reasons**, one type of nucleus (out-gliding or in-gliding or steady-state) comes to be favored for a time in one area or another. (Stockwell and Minkova 1988a: 371 - 2; their boldface)

Thus Stockwell and Minkova reject a general tendency toward raising and thereby reject even the relatively weak claim that the GVS is unified in a pragmatic sense, that is, as a handy way of systematizing vowel changes that are not genetically related to one another. In fact, Stockwell and Minkova cite dialect evidence which shows that lowering may occur, for instance in "the lowering of the high diphthongs, by the lowering of [u] to [Λ] or even [a], by the lowering of Middle English o go [α] in many dialects, by the development of [⊃i] in Australian, etc. The two **constant** tendencies are for diphthongs to dissimilate and, having dissimilated maximally, to reassimilate" (1988a: 373; their boldface). In their closing section on the structural coherence issue, Stockwell and Minkova conclude that there is none, that is, that the sound changes traditionally seen as part of the GVS are actually independent and do not form part of a unified phenomenon.

Stockwell and Minkova's conclusion is paradoxical given their commitment to the four Germanic tendencies listed above. Contrary to Stockwell and Minkova's implication, unity among sound shifts does not require uni-directionality. It does require some shared trait, but arguably that commonality can be provided by any consistent motive principle or principles. So what if dialect evidence shows a more complex pattern of motion than raising? That fact (if Stockwell and Minkova are right in asserting that it is the case) could as easily motivate us to change our view of what the GVS is rather than to declare the GVS a fantasy. We could say, for instance, that Jespersen's famous 1909 diagram shows just one instantiation of the four principles of Germanic sound change which Stockwell and Minkova propose. That would not be to deny that the GVS existed as a unified phenomenon nor that it was *sui generis* in some sense any more than to recognize the principles of classical mechanics is to deny the unity and typological uniqueness of, say, the class of all falling objects in general or of descending snowflakes in particular.

Stockwell and Minkova's boldfaced qualification, "**for social but not linguistic reasons,**" deserves special notice. It tells us that the authors locate some principles of linguistic change, including phonetic evolutions, in the realm of physical language production and reception; they place other changes in the realm of social history, making them accidental with respect to language *qua* self-sufficient "machine." What Stockwell and Minkova present, then, are two distinct realms, language and social circumstance. The principles of language change, if perceived as efficient causes, can reside in either realm. (At present we are interested only in phenomena within the language, but presumably language can affect social circumstances as well.) Because language and social circumstances are perceived as two distinct, discontinuous contexts in Stockwell and Minkova's treatment of the issue, however, it would be difficult to perceive as unified a process which is actuated by efficient causes from both realms. But a final cause of the sort Lass relies on in 1976 is a part of language itself and must reside wholly in that context. Stockwell and Minkova's caveat, "**for social but not linguistic reasons,**" is thus one more indication that the authors' metahistorical commitments stack the deck against unity.

Whether we take the GVS as a phenomenon in the sense of being causally unified or as a means of systematizing data with no implications of causal unity, we encounter disagreement among the interlocutors in the unity debate. As we saw in the case of the causal criterion above, the best we can do is to acknowledge the respective metaphysical foundations of the various positions. Neither understanding of phenomenon -- as that which is causally unified nor that which systematizes diverse data -- is necessary. But both are possible, Stockwell and Minkova's assertion of multi-directionality among vowel shifts notwithstanding.

### (c) "Real" linguistic phenomena: those which are perceived directly or which successfully explain and systematize extant data.

"Real" as an ontological qualifier is a complicated concept, but it is as central to our discussion as "unity" and "phenomenon" and so cannot be ignored. As we will see shortly, the reality of sound systems (not just of individual sounds) plays an important role in Lass's early commitment to the unity of the GVS.

There is another, more oblique way in which Lass emphasizes theory over data and thereby makes theoretical entities "real." Moreover, he maintains this perspective even in his 1992 paper, after he has made significant concessions to Stockwell and Minkova's empiricist badgering. Lass frequently speaks of the creative aspect of history (and by implication, of historical linguistics in particular). What he wants to emphasize is that there is more to linguistics than perception; the theoretician *creates* linguistic history, at least to some extent. Historians, says Lass, "make history out of phenomena." This is consistent with the notion that the subject matter of history is only partly independent of the observer. The historian may end up creating more than she observes. "Like many historical issues, those surrounding the GVS are not really empirical; some are logical or methodological, others philosophical" (1992: 145). Finally, Lass claims to be doing "metahistory" in his 1992 paper. We saw that in the case of the Constellation Fallacy Lass views "creation" as misguided, but in other contexts he sees the creative impulse of phonologists as productive and inevitable.

The pre-1992 Lass is prepared to say that the GVS is recognizable as a phenomenon only in hindsight (the argument from final cause). At the same time, he is anxious to defend the scientific character of his commitment to phonetic systems and of theoretical generalizations about groups of phonetic data. That means that propositions about the GVS must be empirically testable (arguably the *sine qua non* of science if we update "testable" to Popper's "falsifiable"). Thus his defense of testability and therefore of the scientific character of his perspective: "Eventually I want to claim that there are extensive classes of sound change which involve 'generalizations' if you will larger than those statable either as single rules or as schemata. And further, that these generalizations are 'real' ones, justifiable by testing as reliable as any we can get in historical matters. These generalizations are at least as real as, and possibly more so than, alternative ones, based on different metatheoretical notions" (Lass 1976: 56).

Lass's mention of testing emphasizes two points. First, we notice again the respect paid to theory, which is what Lass's "generalizations" amount to: they are real, but in a theoretical sense. That raises once more the issue of truth criteria, namely that we may have to test for coherence in the realm of theory rather than for a correspondence between theory and an objective reality. This is especially true given Lass's acceptance of sound *systems* (rather than just individual sounds) as primitives in his analysis of the GVS.

Lass's 1969 emphasis on the coherence of linguistic theories as a measure of their truth and his obvious approval of Stockwell's metaphor of "intricate webs of suppositions and inferences" do not in themselves tell us what Lass takes to be the units of phonetic theory. This is a crucial question in the unity debate, since whatever the criteria of unity, we need to know what sorts of things can be unified with one another. For instance, if we conceive of unity as stemming from a shared cause and a compact temporal context, we need to know what things can share causes and between what things temporal boundary lines are to be drawn. It is tempting either not to question the nature of such units, or if the question is posed, to answer immediately: *individual sounds* are the units which phonology studies. They are the natural atoms, that is, things which cannot be or are not to be further divided. They are the natural focal points, in the sense Euclid meant when he said that "a point is that which has no part" (153), that is, which cannot be divided into any more basic, more primitive parts.

But in 1976 Lass offers us a different answer. "...[P]honological systems *per se* (i.e. structured inventories of segments in opposition) are necessary primitives in phonological theory" (Lass 1976: 51). It seems clear that Lass means by the word "primitive" the atoms or Euclidean points of his version of structuralism. By announcing that phonetic systems are primitives, Lass makes it clear that he is going to treat such systems as *real* entities instead of mere mental constructs imposed on collections of individual phonemes which can be physically sensed, measured and described. A rigorous definition of "system" need not occupy us here. For present purposes it is enough to remark that a set of data such as those traditionally subsumed under the GVS can be considered a system so long as we can demonstrate or even suspect the existence of a unifying attribute.

It seems natural to respond by asking for a justification of this commitment to systems of individual sounds as primitives rather than to the individual sounds themselves. But in 1976 Lass seemed to feel little obligation to justify the status he gives these systems. The fact that he treats them as primitives, however, is another indication of his emphasis on theory and inference over data which are merely perceived. It could even be argued that the boundary line between theory and data necessarily begins to blur when sound systems are declared to be primitives. The unity necessary to make a system out of individual sounds must be real in some sense, but we recognize the reality not by perceiving it directly but rather by creating it in our theories and then "perceiving" it there. Moreover, for Lass it does not detract from the primitive character nor from the reality of sound systems that we recognize them in hindsight. Again he asserts without justifying, apparently relying on his own and his readers' linguistic intuition: "We feel constrained to say, not that events A, B, C occurred and we can conveniently call them 'the X'; but rather 'the X' occurred, and the stages in it were A, B, C" (Lass 1976: 53). But surely there must be constraints on such judgments: there cannot be a blanket warrant justifying assertions of unity with respect to any arbitrarily selected collection of phenomena. What are the considerations which constrain us? Certainly the Lass of 1976 would not see empirical "facts" as severely limiting the range of possible unitary phenomena. In the middle of his 1976 chapter on the GVS, he goes so far as to use the phrase "holistic intuition" in describing the phonologist's means of approaching a phenomenon such as the GVS (Lass 1976: 68).

To summarize what we have to this point: the Lass of 1976 is committed to the reality of sound systems, placing them on a par with individual sounds in so far as their actuality is concerned. Further, he is not bothered by the fact that these sounds are perceived only in hindsight. Finally, he takes this position intuitively, without offering a rigorous argument for it. From this point it is no leap and hardly even a step to perceive the GVS as a unified system. That the sounds are not static should not prove much of a barrier to something as potent as "holistic intuition." (It appears that the phrase could be applied to Lass's gut feeling that the GVS is a single, coherent phenomenon. Nonetheless, he quickly adds that this intuition is, at least in some of the arguments which accompany it, "firmly based on dialect evidence" (Lass 1976: 68). Whether based on the phonetic facts of dialects or not, these arguments do more than simply point and say "Look there! Q.E.D.!" Instead, the arguments are themselves part of a theory about the data.)

Lass's defense of sound systems rather than just sounds as the units of phonetic structure confronts us with a second question: Can we test propositions about systems themselves in correspondence fashion or only about the relationships between systems, that is, in respect to their coherence? Presumably there are theoretical criteria which determine what collections of data qualify as coherent systems. But in 1976 Lass does not address the particulars of recognizing phonetic systems. Instead, his presentation remains intuitive and philosophical. He hasn't shown us a way to demonstrate *empirical* unity in shifts of English long vowels and diphthongs. In fact, he seems content to argue philosophically, using empirical data more to illustrate than to prove. He does use rhetorical tactics, such as asserting that linguistic consensus is on his side. "Jespersen and Luick agree on one fundamental point: the shift is a single event, with an inner coherence" (Lass 1976: 58). On the other hand, he senses that certain descriptions of the GVS *qua* system misleadingly underemphasize its unity. He describes Chomsky and Halle's 1968 formalisms as "not apparently related to each other, nor necessarily directly adjacent either historically or synchronically, nor as far as we can tell anything but adventitious" and then quickly reiterates that orthodoxy supports his view: "But scholars have unhesitatingly identified the GVS as having an 'inner coherence.'"(Lass 1976: 64). Once again, it is theory rather than data, only this time with a little help from scholarly consensus, which determines whether unity exists or not.

Consistent with his attention to theory and to a coherence understanding of truth, Lass sought to broaden the terminology used to debate the unity question. He found the received ontology of structuralism to be too restrictive. If we acknowledge the reality of things like tongue heights and sounds (frequencies), we still do not possess sufficient building blocks to explain what had happened in the history of shifts among English long vowels. This process of ontological enlightenment seems to have had three elements for Lass.

a. Willingness to believe that our descriptions of sound changes -- often taking the form of rules -- are real, not just formal. Thus he asks rhetorically: "In particular, how can we locate the impulse as raising, and yet keep this separate, in a sense, from the rules that effect it?" (Lass 1976: 68).

b. The same goes for the concepts employed in these descriptions, i.e., we must enlarge our ontological vocabulary: "Clearly, in cases like the GVS, it will be necessary to include, for explanatory purposes, notions like 'phonological space' and 'chain'. In other words, things like inventories and paradigms, etc., may have to be given more than the epiphenomenal status they now have in standard theory..." (Lass 1976: 68 - 69).

c. A willingness to look at what we could call macroscopic, not just microscopic, changes. Lass calls the broad-brush changes "metarules," which are to be distinguished from the simple rules describing local changes. In the case of the GVS, Lass wants us to understand that the overall tendency of individual sound changes, a tendency which he terms "raising," is as real as the individual sound changes themselves.[69]

What emerges from Lass's 1976 musings is a two-tiered reality. The fact of the matter necessarily includes phonological data, but it can only be expressed with the aid of theoretical constructs such as "raising." Which entities are to be taken as real and which as "epiphenomenal," to repeat Lass's term, is at least to some degree a matter of choice. And the choice a given linguist makes is dictated by the overall theoretical context in which he works.

Perhaps it would be most accurate to treat reality as a quality which is attributed to linguistic entities on the basis of the linguist's intuition rather than through careful analysis and argument. On this view the linguist would have no obligation to explain why a given phonetic entity (e.g., individual sound or sound system) is to be considered real. Something like this reliance on intuition seems evident in Lass's casual treatment of sound systems: he offers examples showing how they function but no arguments justifying his belief that they are real as "mirrors" of an extra-theoretical world rather than merely useful as supporting members within linguistic theories.

But appeals to intuition may be unsatisfying when they occur in the middle of didactic arguments, depending on one's metaphysical tastes. Even if we decided to trust Lass's intuition, we still would not know what reality means in the case of sound systems such as the ones he presents for our consideration in 1976. Does his linguist's intuition somehow allow him to *perceive* such sound systems directly or does intuition lead him to *create* them in the process of theorizing? I suggest we satisfy ourselves for the moment with a prescriptive rather than a descriptive approach to the problem. Rather than ask what Lass meant by treating sound systems as real entities, then, we will simply say that he *should* have meant something like the following. Phonetic entities are real (1) if they are perceived directly or (2) if statements about them can be tested empirically. The first condition could be taken to embrace not just sounds that are (1a) heard by nearly all subjects and attributed to a definite cause external to the hearers, but also those that are (1b) imagined (usually by a single individual) or (1c) caricatured. Someone whose ears are ringing calls the sound real in the sense of (1b) even if the sound lacks an external cause identifiable by most or all other listeners; similarly, a linguist's orthographic scribblings represent real sounds (in the sense of 1c) even if the orthography itself is only symbolic and approximate.

Obviously these criteria leave huge latitude for attributing reality to linguistic phenomena. Senses (1b), (1c) and (2) could all be wielded to argue the reality of the neighing of unicorns. But it should be remembered that our immediate goal is to find out how to interpret claims for the reality of linguistic entities such as sound systems. For that purpose, it is reasonable to start with a very broad brush and then progressively refine our understanding of what is meant.

It is tempting to understand Lass's application of the term "primitive" to sound systems as a claim that these are somehow directly perceived in one of the senses of condition (1). If (1a) is too problematic because of the disagreement among linguists on the unity question, then Lass can fall back on (1b): he and some linguists perceive these systems (through "holistic intuition, perhaps), even if others do not, and thus the systems are real to them as perceivers. If that position is attacked as narcissistic and unscientific, then there remains (1c): granted that the traditional GVS as a system is imperfect in that there are pronunciations (e.g., *great, break, steak*) and possibly whole dialects which are ill accommodated by its rules, the system is nonetheless descriptively powerful in most instances. Therefore, it is real. This last argument, (1c), takes us close to the second condition -- empirical testability as the acid test of reality.

It is harder but nonetheless possible to view Lass's sound systems as fulfilling condition (2): harder because we cannot test propositions which require interviewing Shakespeare or Cromwell, possible because sound systems can figure in hypotheses which are testable to the degree that they account for the extant data, including texts, orthoepic testimony and present pronunciations, either well or poorly. In other words, a particular sound system might be seen as real in so far as it figures as part of a hypothesis which does a good job at something. That something might be justifying a proposed cause for or systematizing a body of empirical data. But before we can deal with the assumption that phenomena can be empirically tested for unity, we need to know something more. First, we need to know what it means for a proposition to be tested, that is, to "fit" the data, and whether such tests can mow down the field of candidate theories until only one remains. Secondly, we need to know how to phrase a proposition about reality. This second question will tend to take us on an elliptical path, rhetorically speaking. That is because when we ask what "things" or "entities" are candidates for reality we risk taking the fact that a reference exists as indication that its referent exists, thus begging the question of whether the thing in question is real or not. We must also be careful not to beg the unity question itself, since when we identify a candidate for reality -- e.g., a sound system as opposed to an individual sound -- we tacitly say that the candidate is unified in so far as it can be named as though it were a single thing. We will look more closely at these issues in the next section.

### (d) Explaining Lass's and Stockwell's trajectories

Probably it is impossible to say with certainty why Lass and Stockwell seemed to agree in 1969, then diverged over the ensuing twenty years, and then grew somewhat closer again after 1992. Although their 1969 positions were doubtless influenced by many sources, it seems clear that they were interested in issues which had gotten a lot of press in the previous years thanks in large measure to the philosopher of science Thomas Kuhn, the philosopher of language Gilbert Ryle, and the philosopher and logician Willard van Orman Quine. None of these gentlemen appears in Lass's or Stockwell's 1969 bibliographies, but their ideas were so ubiquitous that the influence can be safely assumed. Even if these philosophers were not a direct influence, we can use them as a means of highlighting key ideas in Stockwell's and Lass's early and late positions.

It is interesting to note that early in their careers Stockwell seems to have been more ally than opponent to Lass on one score. Both seem to have shared a common conviction that empirical fact is a necessary but insufficient basis of generalization in historical phonology. "...[H]istorical 'facts' about pronunciation (the kind that fill our historical grammars) are virtually meaningless outside an interpretation imposed upon them by a theory" (Stockwell 1969: 243). In his introduction, Lass quotes Stockwell with obvious approbation (Lass 1969: 3).

But how can one hold Stockwell's belief in the primacy of theory and yet deny the unity of the GVS? If we *cannot* logically conceive of the GVS as a single unified phenomenon -- Stockwell's position (with Minkova) nineteen years later -- then assuming the primacy of theory over fact would mean that there is no coherent theory which allows unity. That seems patently untrue, as evidenced by Lass's reading of Martinet, for example. What must have happened is that Stockwell gave up his belief in commitment to theory and moved towards a balder empiricism while Lass stayed firmly put...until 1992. Even then, Lass harbored reservations which allowed him to find a core of unity within the traditional GVS.

Why the change of opinion or, to put the question more broadly, of emphasis, first on Stockwell's part and later on Lass's? How can we characterize the currents which pulled Lass and Stockwell away from one another? Speculation is not idle here, since by uncovering possible outside motivations we may see more clearly how the players treat our three assumptions.

It seems conceivable that in the late 1960s Lass and Stockwell were influenced by Thomas Kuhn (1962). Besides the dialectical model of scientific development which Kuhn proposed and which may or may not be applicable to linguistics itself (q.v. Greene 1974, esp. 499), the message which many took from Kuhn applied to the relationship between facts and theories within science: facts alone cannot provide scientific meaning; science emerges from subjective theory and culture; “facts” and observations are not “objective,” since they are influenced by theory and culture. The question of how theories are developed is central here. Kuhn argued that science does not begin with a neutral collection and analysis of data culminating in the formation of explanatory and predictive theories. Rather, he suggested, the scientific enterprise is always colored by social factors and cultural commitments. The notion of theory-ladenness is clearly echoed in Stockdale’s phrase (quoted above): “...‘facts’... are virtually meaningless outside an interpretation imposed upon them by theory.” This is consistent with Kuhnian normal science in so far as Stockwell holds in 1969 that the interpretation of data depends upon a background of commitments which are not strictly empirical. But it is doubtful that the early Stockwell is such a thoroughgoing relativist in this regard as Kuhn. For both, however, theory plays the primary role, providing the context which alone allows facts to be interpreted in some meaningful way.

As we have seen, Stockwell and Minkova’s use of the title “English Vowel Shift” in place of the traditional “Great Vowel Shift” indicates their sense that the question of phenomenal unity has been begged by a misleading phrase, among other factors. The essence of this worry was perhaps most thoroughly discussed by Gilbert Ryle in 1931:

> But the search for paraphrases which shall be more swiftly intelligible to a given audience or more idiomatic or stylish or more grammatically or etymologically correct is merely applied lexicography or philology -- it is not philosophy.
>
> We ought then to face the question: Is there such a thing as analyzing or clarifying the meaning of the expressions which people use, except in the sense of substituting philologically better expressions for philologically worse ones? (Ryle 1931: 86 in Rorty 1967).

Ryle believes that users of what he calls "Quasi-ontological Statements" in fact know very well what they *mean*, and what they *mean* is logically consistent. But what they *say* is syntactically confusing, and hearers may be misled, which is to say that they may understand something which was not meant (Ryle 1931: 87 - 89 in Rorty 1967). Ryle’s prescription for eliminating existing and potential confusion of this kind is to avoid quasi-ontological statements through careful rephrasing. Thus it would be careless to say that “Unicorns are white,” since predicating Y of X implies to some that X exists, whether X is in fact real or not. Far better, advises Ryle, to ensure that receivers as well as senders of linguistic information are not misled. This is accomplished by replacing statements such as "Unicorns are white" with propositions like, "There are horse-like animals with a horn protruding from their foreheads and such animals are white." A statement of this form speaks explicitly of existence as well as whiteness. In so far as existence is made to look like an explicit predicate (not to push aside a centuries-old philosophical debate as to whether existence can actually be predicated), our attention is turned toward the question, Are there really such things as unicorns? But there is no promise that we will find sufficient data to answer the question.

By contrast, Stockwell and Minkova are convinced that many linguists who speak of the Great Vowel Shift are themselves deluded. But these misguided phonologists are not beyond help, and Stockwell and Minkova believe that the remedy is precisely what Dr. Ryle had prescribed. By speaking of the Great Vowel Shift *as though* it existed, Stockwell and Minkova argue, linguists fall into the trap of believing that it *did* in fact exist as an historical entity. Thus the spell can be undone by using phrases such as the English Vowel Shift, terminology which does not beg the question of existence before analysis has begun (or so Stockwell and Minkova believe).

But at this point we should ask whether the question of the GVS's existence as a unified phenomenon can be settled at all. One approach to the question is to pay attention to the state of our current empirical knowledge as opposed to what we may someday know. This distinction is often made with respect to linguistic questions. "It must be emphasized...that the principle that there are no primitive languages is not so much an empirical finding of linguistic research as a working hypothesis. We must allow for the possibility that languages do differ in grammatical complexity and that these differences have not so far been discovered by linguists. It would be as unscientific to deny that this possibility exists as it is to say that Latin is intrinsically nobler or more expressive than Hottentot or one of the Australian Aboriginal languages" (Lyons 1981: 30 - 31). Consistent with this outlook, the mature Lass cautions that his slimmed-down version of the GVS is based on "corrigible" conclusions (Lass 1992: 152). By this he seems to mean that new empirical discoveries may necessitate new conclusions. But there is another sense in which conclusions may be corrigible independent of what happens on the empirical front, a sense that theories are inevitably tentative and that no single theory accounts for any given body of empirical data. That view evokes one figure among contemporary philosophers perhaps more surely than any other -- Quine, whose Indeterminacy of Translation Thesis we have already considered above.

Quine is well-known for his advocacy of the role of inference in theory-building. His famous essay "Two Dogmas of Empiricism" (1951) appeared the year before Stockwell completed his dissertation (cf. Stockwell 1969: 243, n. 3). The work is seen as one of the signature documents in the backlash against logical positivism (sometimes also called "logical empiricism; q.v. Rosenberg 1993: 64). Quine rejected the positivists' distinction between analytic and synthetic truths (first dogma) as well as their contention that all meaningful statements are equivalent to propositions about immediate experience (second dogma). While freeing himself from the first dogma (and from the influence of his teacher Rudolf Carnap), Quine defended a view encapsulated in the phrase "web of knowledge," meaning a kind of knowledge which is neither experience nor theory, exclusively. Quine sees knowledge as being woven from experience (data) *and* theory, among other strands, and he emphasizes the difficulty of separating the two ingredients. Stockwell's 1969 phrase "webs of suppositions and inferences" may be coincidental in its use of the same metaphor, but its substance can be read as Quinian. In rejecting the second dogma Quine emphasizes the impossibility of knowing merely by observing; knowledge presupposes theorizing as well.

What is not clear as of 1969 is whether Stockwell endorses the whole of Quine's attack on logical positivism. Perhaps the Vienna Circle's best known claim is that metaphysical and theological statements are meaningless in that they are neither empirically verifiable nor (as many claims of mathematics are) analytic. It seems clear that the body of theory associated with the GVS is not a collection of analytic statements. But does Stockwell think phonological theories are empirically verifiable, and if not, will he reject them as nonsense (as a card-carrying positivist would have done)? A close reading of Stockwell (1969) leaves that possibility open. Lass, on the other hand, is more explicit in his rejection of empiricism in general and of its twentieth-century instantiation in positivism in particular.

The ITT is clearly relevant to the GVS unity debate, but again it is unclear whether the Stockwell of 1969 is prepared to say that a theory may be valid in some sense even though it is not wholly determined by the data. The notion of knowledge as a "web" of observation and inference can be wielded by those who insist a valid inference have no competitors as well as by those who accept that absolute decision criteria may always be lacking. This tension applies to higher-level questions than, say, whether the GVS was actuated by a push-chain, a drag-chain, or by some other efficient mechanism. A hard-core neo-positivist would have insisted that the unity question itself is metaphysical nonsense so long as it cannot be conclusively tested in an empirical context. Stockwell might always have belonged in that camp, but in 1969 it was too early to tell. At that time he had not dealt with the problem of whether the extant empirical data relevant to questions such as the unity debate must uniquely determine a theory before it can be accepted. He certainly had not tackled the specific question of whether the GVS is a single, unified phenomenon in the way he and Minkova would do in 1988.

We have already seen a quotation from Stockwell and Minkova (1988a) expressing their belief that psychological factors may have motivated supporters of the traditional GVS. The general response to Quine suggests that Stockwell and Minkova themselves may have been under some psychological pressure to adopt their 1988 position. Although Quine found supporters, he also alarmed thinkers in various disciplines because his arguments seemed to bolster the R-word -- relativism. If data underdetermine theory in the way and to the extent Quine suggests, then truths of all kinds, even scientific ones, must be seen as much less certain than many would like to believe. There is always the possibility that an alternative, quite different theory will match the extant facts just as well as the one currently in vogue among scientists. Some of Quine's detractors would have resented being labeled "empiricist," but it is nonetheless true that these opponents tried hard to shore up the reputation of empirical data, a reputation which they felt Quine had tarnished. Often these defenders of the primacy of data over theory took a moderate tack, merely insisting that even if such information is not sufficient to determine a single, correct theory, it is nonetheless a powerful tool in paring the list of candidate theories. Through this measured approach the negative role of empirical data came to be emphasized: it would be the iconoclast, it would shatter the ambitions of theoreticians.

In short, Quine attacked the positivists at their weak point: dealing with non-observational (i.e., theoretical) elements in human knowledge. Rosenberg summarizes this weakness nicely:

> Twentieth-century empiricism has had great difficulty reconciling its claims about empirical meaningfulness with the apparent commitment of scientific theory to the existence of entities beyond our observational access. Empiricists since Hume have either sought to translate claims about theoretical entities into statements about what we can observe or sought to treat such claims as convenient instruments or heuristic devices with only apparent semantic content. (1993: 67 - 68)

But the extent of Quine's relativism shocked many to the point that they began to move back toward what they hoped would be a new, more rational empiricism. Did Stockwell take part in this migration between 1969 and 1988?

As of 1969 it is still unclear who owns Stockwell's soul: he's not a full-blown positivist (the movement was passé even before he hit graduate school), but it is not clear whether he leans toward Quinian relativism or in some other direction, perhaps a milder empiricism than that associated with the Vienna Circle.[70] One can read the early Stockwell as adapting to historical phonology some of Quine's attacks on logical positivism, but as of 1969 we do not know whether he will be willing to follow Quine down the road to thoroughgoing relativism of the kind many critics read in the ITT. All that is certain is that the early Stockwell sees knowledge of historical linguistics as being dependent upon theory as well as data. Some weak version of the indeterminacy thesis seems inescapable even though Stockwell does not address it: more than one theory will account for most bodies of empirical data. But does that mean we should accept the stronger thesis that there is no reason to prefer one competitor theory over another? Does it mean that we should commit to no theory without authoritative (i.e., wholly empirical) verification?

Stockwell (together with Minkova) shows his cards in 1988. I hesitate to call him an empiricist. That label has been applied to such a wide range of quite varied positions that it is not always clear what is meant. Perhaps we can call him a negativist. The basic pattern of his and Minkova's argument is this: If we look at all the extant data -- including texts, orthoepic testimony, and information gleaned from contemporary dialects -- we can indeed infer some basic rules of vowel changes. In other words, we can construct generalizations which fit the data. But what we cannot do is generalize the generalizations, to construct a super-phenomenon such as "vowel raising." In reality some English vowels have lowered. Thus if there is no way of identifying an overall pattern such as raising, it is certainly vain to discuss macro-mechanisms such as push- and drag-chains in order to account for an overall pattern which does not exist.

More abstractly, Stockwell and Minkova's argument begins by using empirical data positively -- to inductively fashion generalizations describing how vowels change. At some point the data are exhausted; there is nothing more to be said, since further theorizing would have no empirical warrant. Confronted with the claim that shifts among long vowels fit the GVS pattern, Stockwell and Minkova's gripes fall into two categories, counter-examples and empirical groundlessness. They can produce counter-examples showing vowels which seem to behave contrary to the GVS pattern, and they claim they can show that there is insufficient evidence to warrant any argument for GVS unity.

Lass, by contrast, can be read as remaining consistently Quinian right up to the appearance of the 1992 paper in which he gives up part of the traditional GVS. His enthusiasm for theory in general is clear in 1969. Shortly thereafter he takes up the GVS unity debate itself (1976), clearly under the influence of a long tradition of believers in unity. In 1986 Lass suggests that Jespersen coined the term "Great Vowel Shift" in 1909 but that "[t]he thinghood of the GVS [was] complete" only when Luick wrote about "*die Große Vokalverschiebung*" around 1929. Luick didn't attribute this phrase to anyone, but Lass believes it was a translation of Jespersen's terminology (Lass 1986: 30). Beyond the evidence Lass cites, one can find indications that the GVS was viewed as a single phenomenon even before Jespersen. Stockwell and Minkova note that in 1885 Holthaus discussed the motive force which began the series of shifts later called the GVS (Stockwell and Minkova 1988a: 356). For the moment it can remain an open question whether such an impulse is necessary or sufficient to justify calling the changes it begins a single phenomenon. What is worth noting here is that as early as 1885 the individual transformations later subsumed under the GVS were considered as a group, but that this was possible only by inferring a cause, that is, by theorizing rather than simply perceiving. In particularly, external impulses such as the one Holthaus sought in 1885 or internal impulses such as those evident inside chain reactions emerge only within the context of a theory.

This brief history of the GVS as concept is consistent with two arguments. First there is a rationale for asserting the unity of the GVS based on the way it is perceived. This argument is reminiscent of the discussion of the real as that which is taken as actual by an individual observers or group of observers (cf. section (3)(c) above). Second, Stockwell and Minkova's discussion of Holthaus (1885) makes it clear that long before Lass (1976) or even Martinet, an outside efficient cause could be seen as sufficient grounds to assert unity (a conclusion Stockwell and Minkova would reject).

But in 1969 it is uncertain how Lass and Stockwell react to such a position. At that time both emphasized theoretical considerations and seemed to accept a coherence sense of truth. Lass shows no sign of questioning the primacy of theory and inference. Stockwell, on the other hand, distinguishes between two types of historical phonology. The first he describes as being "oriented to the physiological facts of articulation, to speech, *parôle*, in a concrete form"; the second as "oriented to the structure, the system, the *langue*, of speech production and perception." He calls the first theory "concrete" for short and the second "abstract" (Stockwell 1969: 229). For our purposes it is of special interest to note that under the "concrete" category Stockwell places "the works of Jespersen, Wyld, Luick, Ellis, Zachrisson, Dobson, Kökeritz, Orton, Kurath," while for him the abstract school is represented by "Trnka, Vachek, Martinet, Lamberts, Reszkiewicz, and most recently Halle and Keyser" (ibid.). But he did not make much more out of this distinction; certainly he does not apply it to the unity debate in a clear-cut way. Almost twenty years later he and Minkova (1988a, b) divorce themselves from thinkers in both the abstract and concrete camps, while Lass finds common ground with both groups, even after 1992.

On one level the arguments offered in this chapter amount to the simple assertion that conclusions depend on assumptions. That is a weak and uncontroversial thesis. But it is less clear *which* assumptions form the metalinguistic linchpin of Lass's early and late positions on the unity question, and the same goes for the viewpoints of Stockwell and later of Stockwell and Minkova. The three assumptions (A1 - A3) proposed above do not cover the entire metalinguistic background of these students of the GVS, of course, but the assumptions go a long way toward clarifying how camouflaged changes in "meta-commitments" affect the more explicit rhetoric.

The first assumption (A1) pointed out various possible understandings of the causal criterion. It would be unrealistic to demand that each term relevant to the unity debate be carefully "unpacked" in every treatment of the subject. On the other hand, dissension in the unity arena is partly a product not just of differing semantic commitments but also of failure to recognize that those semantic differences exist. We have seen that the term "unity" itself is bound up with causal and temporal criteria. But a close reading of the pre-1992 Lass and his opponents makes it clear that the two sides do not adopt the same concept of causality. Lass's hyperbolic trajectory, in which he begins by accepting the entire traditional GVS as a single phenomenon and later decreases his commitment by more than half, can be explained in large measure by his abandonment of final and outside efficient cause in favor of inside efficient causality. The significance of Lass's early position is that it allowed him to speak of an impetus efficiently motivating the entire GVS, as a unit, as well as of agents for proximate changes among individual vowels and within distinct time frames. This view stands in sharp contrast to that of Stockwell and Minkova (1988a, b), who reduce the impetus of vowel changes to inside efficient causes in the form of four Germanic tendencies which take place in a proximate context. That Stockwell and Minkova lack any sense of a final or an outside (non-participating) efficient cause as operative in the GVS is perhaps best exemplified by the brevity of their treatment of the inception problem. They ultimately dismiss the matter as a "pseudo-problem" (Stockwell and Minkova 1988a: 357). Moreover, it seems that they do not recognize the huge role played in this verdict by their "meta-commitment" in favor of inner efficient cause and against outer efficient and final causes. By contrast, the Lass of 1976 shows signs of recognizing his own semantic assumptions about causality (and therefore about unity). This is particularly evident in his explicit mention of final causality.

The Lass of 1992 seems to have given up final causality altogether, but it is not immediately clear whether his taxonomy of efficient causality has eroded to the point where outer efficient causes are excluded as well. One indication of his changing metalinguistics in this regard is his discussion of what he calls the Constellation and Zebra Fallacies. The Constellation Fallacy seems to apply only to a synchronic questions of the form: Does X "really" look like Y or is the resemblance all in someone's head? The Zebra Fallacy, on the other hand, identifies the mistaken perception of a genetic (and necessarily diachronic) relationship where in fact a more superficial, cosmetic similarity is all that exists. For the earlier Lass the GVS unity debate is patently diachronic in that the key issue is whether the individual sound shifts are related to one another or not. That implies that interlocutors in the debate are at risk of committing the Zebra Fallacy, but it is difficult to imagine a case in which a phonologist would err in his evaluation of the GVS's unity in a way consistent with the Constellation Fallacy.

Why, then, did constellations engage Lass in 1992? One possible answer is that, either worn-down or star-struck by the empiricism of Stockwell and Minkova (1988a), he had begun to de-emphasize the causal criterion of unity in favor of some other condition such as the temporal one. That would not mean, of course, that he need abandon causal analysis entirely. On the contrary, his 1992 argument makes it clear that the causal relationship among individual vowel shifts is crucial to the unity of his new, slimmed-down GVS. But it is noteworthy that these causal relationships are all efficient rather than final, and moreover, that they are all inner (participating) efficient rather than outer (non-participating) efficient. In other words, Lass has abandoned his sense that the GVS traces the journey toward a pre-determined *telos* or end-state of English long vowels (the final sense), and he has likewise given up his defense of an outside (non-participating) cause motivating the change of all vowels taken as a pre-transition unit (as, for instance in Martinet's symmetry principle acting on the *isochronie* which embraced English long vowels collectively, i.e., as a unitary system). Inner (participating) efficient causes can be evident in shorter time spans than final or outer (non-participating) efficient causes. This is particularly true when phonetic change occurs through the mechanism of dialect borrowing, the explanation which Stockwell and Minkova prefer.[71] But even where dialect borrowing is not seen as the mechanism of phonetic change, and even where sound changes are believed to have occurred over extended time periods, a "snapshot" view may suffice to analyze a given causal relationship. This is where Lass's Constellation Fallacy comes into its own.

The second assumption deals with temporal gaps -- the issue of whether the apparent cessation of a motive principle necessarily marks the boundary point of a phenomenon. Put in the form of a question: Can a unified phenomenon be animated by a principle which seems to act sporadically, or must any apparent cessation of agency be taken as proof that the "phenomenon" in question is actually a disunified collection of smaller, independent sub-phenomena?

There is no *a priori* reason why an ostensibly staccato agency must be inconsistent with unity. First there is the possibility that a motive force may in fact act continuously even though its uninterrupted character is not evident in its effects. That should be obvious given the robust and generally accepted theory of Darwinian evolution, in which the mechanism of natural selection through inherited characteristics is always at work, even though some empirical evidence such as the fossil record shows long periods of stasis broken by intermittent episodes of comparatively radical change. One currently popular but controversial variant of evolution is the theory of punctuated equilibrium, which seeks to explain how such empirical evidence is consistent with the essential thesis of Darwinian evolution. Punctuated equilibrium is in short an alternative to the notion that Darwinian evolution necessitates continuous gradual change.

Lass (1992) suggests that changes at the half-open level and below cannot be part of the same phonetic phenomenon as the higher changes because there is a temporal discontinuity separating the two "halves." Stockwell and Minkova (1988a) offer an even more explicit attack on unity based on a temporal gap. It is ironic that Lass coins the title of the article in which he endorses Stockwell and Minkova on this score from Stephen Jay Gould, the Harvard paleontologist who is the co-founder and currently the best-known champion of the theory of punctuated equilibrium.

If we look at the temporal issue in the abstract, it seems clear that sequentiality is more important to the unity debate than continuity. Using Darwinian evolution as a model, particularly its punctuated equilibrium variant, it is possible to believe that a series of changes can share some causal criterion sufficient to assert unity even though the agency involved may be sporadic (based on its manifestation in effects) and the changes themselves temporally discontinuous. If we take a closer look at Stockwell and Minkova's (1988a) argument (sections 2.2, 3.2), we see that they cannot brook sporadic change simply because they exclude all but proximate agency: with no possibility of appeal to final or outside efficient causality, they can ground phenomenal unity only in an unbroken chain. But this is again the result of their metalinguistic commitments, which themselves are not necessary to a logical appraisal of the unity question. Other commitments are as logically valid in an objective sense while allowing that unity can exist in the presence of apparently sporadic agency. (Whether the agency is in fact sporadic or not we cannot know; we can say only that its effects have a somewhat broken character.) The early Lass and Martinet again serve as cases in point (section 2.4).

The third assumption considers further semantic issues -- possible understandings of what constitutes a "phenomenon" and what conditions must be met for a phenomenon to be "real" -- which form part of the metalinguistic commitments relevant to the unity question. To understand how these terminological matters figure in the unity debate, we reviewed the well-known distinction between correspondence and coherence understandings of truth. It appears that early on in their careers, Lass and Stockwell agreed that intratheoretical coherence played as important a role in linguistic knowledge as the correspondence between theory and data. Stockwell and Minkova (1988a) and Lass (1992) give empirical data virtual "veto power" over any theory. There is nothing wrong with that in itself. But when theory is eviscerated by a starved repertoire of causal types and by the insistence that continuous agency is a prerequisite of unity, then it is very difficult (or perhaps impossible!) to find a theory which is consistent with the extant empirical data and at the same time supports the unity of the traditional GVS. Under these conditions, unity would only be possible if we had a continuous "chain" of data points, every two of which bear a demonstrable causal relationship to one another.

These speculations serve ultimately to call into question what we accept as "real" linguistic phenomena. If a sound or sound change is perceived directly, then presumably we are justified in considering it to be real. But in historical phonology it is obvious that much of what we take to be data is inferred rather than directly observed. We might stipulate that such inferred data are to be taken as real only if they can stand the scrutiny of empirical testing, that is, if they are consistent with directly observed data such as that gleaned from dialect studies.

A phenomenon may also be seen as real to the extent that it effectively systematizes a given body of data. But this is doubtless the weakest sense of the terms "phenomenon" and "real" as applied to the traditional GVS. The implication accompanying this evaluation is clear: the GVS is only a pretty picture, a pedagogical aid for explaining the rough outlines of English phonetics to neophytes, or perhaps a spur to creative investigation of how the language really evolved.

Finally, we spent a few pages trying to understand the philosophical currents which might have motivated the Lass and Stockwell we observe in 1969 -- both eager defenders of theory's role in the knowledge of language (including phonetics) -- to arrive at their mature positions, in which theory seems to have lost ground to observed data. (Stockwell and Minkova are clearly much more empiricist than even the mature Lass.) Quine's relativism, which appealed to many in the late 'fifties and throughout much of the 'sixties as a way of escaping the empiricist excesses of logical positivism, was seen by others as having gone too far. Thus there was a backlash against the kind of thoroughgoing relativism evidenced by the indeterminacy of translation thesis, which calls into question the very possibility that *any* theory can be justly considered the best fit given a body of empirical data. This movement from anti-positivism (characterized by emphasis on the legitimacy of theorizing and of judging truth on the basis of intratheoretical coherence) to anti-relativism (marked by insistence that empirical data do play the decisive role in arbitrating the claims of rival theories) seems to track with Stockwell's and perhaps with Lass's odysseys through the mazes of the unity debate.

Careful consideration of the three assumptions prompts reflection on the difference between taxonomy and orthopedics alluded to in the introduction. To treat the unity debate as an empirical question in the way that Stockwell and Minkova (1988a, b) do is to perform a pigeon-holing exercise within an already established taxonomic context. If the rules of the taxonomy dictate that final and outside efficient causes are to be ignored and that a 300-year temporal gap necessarily imposes a boundary line between taxa, then Stockwell and Minkova have performed their task unimpeachably. Using inside efficient causes alone yields a picture of shifts among English long vowels and diphthongs scored by multiple fractures. Small islands of data obviously belong together, but it is very difficult to understand how or if all the data, taken together, relate to one another. But what if we approach the same picture -- an X-ray, as it were, featuring observed data with little or no inferred data included -- and question whether it is complete? The obvious answer seems to be that it is not, and if it is incomplete we cannot be absolutely certain that, for instance, the displacement of e: with a: is not causally related to the displacement of e: with ɛ:, even though the two seem to be separated by a considerable number of years. Suddenly we are faced with an orthopedic question (in addition to the taxonomic one):

How do we mend the skeleton, adding to it if need be, so that its form represents the real genetic history of long vowel shifts? It is not that we have lost interest in the taxonomic issue -- the question of where a given micro-change belongs -- but rather that we cannot answer because the orthopedic question presupposes the taxonomic one. There is no point in specifying the correct pigeonhole for a data point until we have established taxonomic criteria, which in turn presupposes a sense, perhaps only vaguely rationalized, of what our principles of organization should be. The three assumptions discussed above indicate three such principles. None is necessary, that is, it is possible to defend a logically consistent theory built on other "meta-commitments." The early Lass is a case in point: his pre-1992 positions are as defensible as his present stance.

## 2. A recursive approach to resolving the disagreement over evolutionary unities

Evaluated abstractly, the disagreement over the unity of the Great Vowel Shift is tantamount to a disagreement over evolutionary mechanism. All of the interlocutors put forth hypotheses which, at their core, assume that evolution of a phonetic sort was responsible for the evolution of the phonological phenomena under consideration. This basic position does not exclude the possibility that the vowel sounds were nudged in one direction or another by something analogous to chance occurrences. In fact, it is easy to ready Stockwell and Minkova as being advocates of a kind of phonetic random walk (Futuyma 1986: 11 - 13; Dennett 1995: 126). That is, instead of believing that sound changes are determined by some standard of aptness, Stockwell and Minkova ascribe to a more or less random cycling through physiologically possible changes. But unless this random development is taken to explain *all* of the vowel changes encompassed by the traditional GVS, then there must be an analog of fitness to explain whatever degree of directionality non-random changes display.

That is not to say that a figure like the early Stockwell rules out chance occurrences altogether. It is always possible that a charismatic prince happened to have a speech impediment of some sort, so that he spoke differently than most of his countrymen. In such a circumstance it is possible that a population might change their speech patterns rapidly and more or less unpredictably as they attempt to mimic their revered leader's way of speaking. But by and large such possibilities give way to a model of evolution which treats pronunciations as phenotypes. A kind of natural selection then drives the progression of sound changes; as new sounds are brought into a region by populations of various fitness (influence), the phonetic character of the linguistic community as a whole evolves. Depending on how one reads the data, natural selection thus operates either as a ratchet function, moving the changes in a more or less constant direction (the early Lass's position), or else it simply sets boundaries to a somewhat cyclical progression (Stockwell and Minkova). Either path may occur gradually or in leaps and bounds (analogous to the theory of punctuated equilibrium in evolutionary biology), depending on the theoreticians' beliefs.

In the model thus described, there must be a concept of fitness to explain why certain physiologically possible sounds do not exist at all in the language and why some sounds evolve into others at certain points. It will come as no surprise to the reader that I find a recursive concept of fitness most easily adapted to the particulars of the GVS unity debate. Since it accommodated a theory of punctuated equilibrium in evolutionary biology proper with ease, it can presumably do the same in the phonetic context. Thus it is clear that arguments against unity which are based on acceptance of the second assumption above -- that temporal gaps in the historical record preclude unity-- are easily dismissed under a recursive interpretation of fitness..

The first assumption likewise presents no great difficulties for a recursive understanding of fitness. Causal relations are in the eyes of the beholder in historical subjects, meaning that any such linkages must be inferred based in part on arbitrary theoretical commitments. From the outset it has been argued that recursive fitness is versatile because of its formality, its "content-lessness." In particular, the form

$$\text{fitness}_{\text{time=t}} = f\,(\,\text{fitness}_{\text{time=t-1}}\,).$$

seems wholly immune to assertions that a causal chain has somehow been "broken." For a causal relationship to be severed, there must be some temporal or logical standards of disjunction. How long is too long an interval between two circumstances to allow for the possibility that they are related? For a recursive interpretation of fitness length of time is irrelevant; the only temporal relation contained in the definition of recursive fitness has to do with succession. The same is true with regard to other logical standards of adjacency versus separation: if someone claims that two events cannot be related to one another because they lack a necessary geographical proximity, that is well and good. But such a judgment is a matter of background commitments unrelated to the formal definition of recursive fitness. The notion "function of" can accommodate virtually any degree of geographical separation, from "touching" to "polar opposites."

Finally, the third assumption which plays a role in muddling the GVS unity debate -- that unity must be an empirical question -- finds no support in a recursive understanding of fitness. This interpretation of fitness provides its own, formal brand of unity through the concept of self-definition. So long as a phenomenon can be expressed as a function of one its earlier states, there is no threat to unity. On the contrary, the unity of the phenomenon is of the tightest possible kind if its component parts are conceived as functions of one another.

## 3. Conclusion

The strategy in this dissertation has been to describe the abstract notion of recursion in basic terms and then to argue that the fundamental structure of recursive manipulation is isomorphic with many arguments employing fitness. Further, this way of looking at fitness offers insight into several contemporary issues of interest to philosophers of science as well as practicing biologists. The various arguments to these ends have seldom constituted rigorous proofs that recursive fitness can function exactly as I suggest. Moreover, compared with the depth and technicality of the literature, some relevant aspects of evolutionary biology were necessarily presented in broad-brush fashion or even caricatured. But perhaps it is enough if the dissertation as a whole has helped create a suspicion in the reader's mind that circularity of a healthy, useful kind is to be expected and even applauded in arguments depending on fitness. That this sort of circularity -- namely, recursion -- has not (to my knowledge) been tapped for the purpose of explaining how fitness functions in evolutionary biology is somewhat mysterious. Perhaps we can use the notion of a cultural paradigm and related ideas which have followed in Kuhn's wide wake to speculate that the conceptual repertoire of computer science is only now making itself felt in this particular instance.

Whether recursive fitness ultimately proves a useful concept or not, there is something appealing in the conclusion that we need not flee from all allegations of circularity. Quite the contrary, if the circularity in question proves to be recursive we may take the accusation as a neutral comment or even (depending on what the circularity achieves) as high praise. It may be too much to hope, but perhaps the idea of recursive fitness will catch on and contribute to an even better account of how fitness functions. If so, the relationship might look something like this:

$$\text{fitness}_{\text{new}} = f\,(\ \text{fitness}_{\text{recursive}})$$

# Notes

[1] The primary source here is Aristotle, *Metaphysics* 985b23 - 986a26, 986b4 - 8; cf. Barnes 1987: 208 - 209.

[2] Whitehead suggests:

> Thus in a sense nature is independent of thought. By this statement no metaphysical pronouncement is intended. What I mean is that we can think about nature without thinking about thought. I shall say that then we are thinking 'homogenously' about nature. (1964: 3)

Von Ditfurth argues that the separation between the individual and "the real world" -- a separation to which he assigns a Kantian basis in so far as spatiality and temporality are modes of perception rather than things which are inherent in a realm outside the perceiving subject -- can be overcome by appealing to the collective. The individual cannot "connect" directly with the real world and so cannot adjust to it; but a broadly-conceived taxon, one which rubs elbows with the external world and adapts itself accordingly, can (1982: 159). What is important in this approach for our immediate purposes is that von Ditfurth has in essence found a way of making intention and extension products of his reading of evolutionary theory, rather than doing verbal contortions to make his reading conform to a realist or non-realist account.

See also von Ditfurth 1982: 143.

[3] The question of whether nature is independently mathematical or whether mathematics is a human overlay on an independently existing reality which in itself transcends such categories was of emotional interest long after the Pythagorean brotherhood had ceased to exist. In a chapter entitled "Mathematics and Imposed Reality," Davis and Hersh (1986: 275) note:

> "The original edition of Copernicus' 'De Revolutionibus Orbium Coelestium' (1543) carried a disclaimer -- written by an editor -- to the effect that Copernicus regarded his system of planetary movement as being only mathematically convenient but not true in a fundamental sense. One usually tells this story and adds a word about the shameful behavior of the timorous editor who misrepresented C's intent."

See also the chapter "A Defense of Internal Realism" in Putnam 1990: 30 - 42.

[4] "...I fell by degrees, in the years 1665 and 1666, upon the method of fluxions..." wrote Newton in *Quadratura Curvarum* (1704); cf. Smith 1959: 614.

[5] The notion that meaning is context- rather than unit-based can be translated into might be translated into a contention that meaning is *process*-based, which is to say usage-based (Alston 1963: 243 in Feigl *et al* 1972). Attempting to understand fitness

*per se* necessitates that we follow the term through its scientific context rather than attempting to capture it in a snapshot. A still photo of motion is itself not in motion, although we may recognize indicators of motion -- a blurriness, a progressions of ghostly images moving in a recognizable path, perhaps -- but we do not see motion itself. To perceive motion itself we must move out of the context of the still photo and follow an object through many contexts. This may also be the case with a concept such as fitness. If we take propositions involving fitness as the analogs of still photos, then we have at best indicators of fitness; we do not have an opportunity to study fitness itself.

[6] That observation alone does not suggest the epistemological underpinnings of the argument which will be made. In case that foundation is of interest, here is a very brief summary of my position. I find the notion that observations are theory-laden very easy to accept, but any claim that theories are themselves *merely* products of culture seems to me an impossible thesis. Probably that is because I think any concept of scientific truth must necessarily contain elements of correspondence and not just coherence. The correspondence may be with an internal rather than an unreachable "external" reality, as in my understanding of Putnam's internal realism, but there is some element of correspondence in the truth criteria underlying scientific theories. This means that some theories, coherent though they might be, will fail to correspond with nature even though they are acceptable or even preferable to competitor theories within the parameters of culture. Scientific revolutions do occur, and although a given revolution may track to some extent with the goings-on in its umbrella culture, a new theory supplants a previous one primarily because it fits some set of facts better than the old one. The culture may select the set of facts, but that doesn't change the fact that there is an element of correspondence which is somewhat independent of the culture.

[7] It should be noted that although Dawkins is often viewed as something of a fanatic, he takes pains to point out that the selfish gene schema is simply *a* way of approaching the unit of selection issue, although he clearly thinks it a particularly fruitful way; q.v. $^{2}$1989: viii - x.

[8] Darwin continues:

> –mentioning these proportional numbers, I may give as an instance of the sort of points, **& how vague & futile they often are** which I *attempt* to work out, that reflecting on R. Brown & Hooker's remark, that near identity of proportional number of the great Families, in two countries, shows probably that they were once continuously united, I thought I would calculate the proportions, of, for instance, the *introduced* Compositae in Grt. Britain to **all** the introduced plants, & the result was 10/92 = 1/9.2. In our *aboriginal* or indigenous flora the proportion is 1/10; & in many other cases I found an equally striking correspondence: I then took your manual and worked out the same question; here I found in the Compositae an almost equally striking correspondence, viz 24/206 = 1/8 in the introduced plants, and 223/1798 = 1/8 in the indigenous; but when I came to the other Families, I found the proportions entirely different showing that the coincidences in the British Flora were probably accidental!--

> "You will, I presume, give the proportion of the species to the genera, ie show on an average how many species each genus contains; though I have done this for myself.--" (Burkhardt 1996: 143-144)

Throughout the discussion I assume that Darwin really believed that Gray might think a desire for raw numbers to be superfluous, but another reading of the remark is possible. Bowlby (1990: 323) notes that Darwin often seemed insecure in his early contacts with the botanist. For instance, as Darwin began to expound his as yet unpublished views on the origin of species in the mid-1850s, he proceeded very tentatively. After sketching his "heterodox conclusions" in a July 1856 letter to Gray, he wrote: "I know that this will make you despise me." In short, Darwin seemed to pepper his early correspondence with disclaimers and "feelers" intended to gain reassurance from the American; the remark about the value of raw data as opposed to percentages -- "...you may think this superfluous" -- *can* be read in that light, though I chosen to take the remark as representing Darwin's real suspicion about Gray's outlook.

[9] Most would not consider Lane and Comac's book as representing good science, but their graphic remarks about bone versus cartilage serve our present purpose well.

[10] This is Kitchener's reading. Actually Charig and Milner are circumspect, merely posing a question rather than fighting tooth and nail for the claw-fishing role of *Baryonyx*: "could Baryonyx have used the claw, like a grizzly bear, for 'gaffing' fish out of the water?" (Charig and Milner 1986: 361).

[11] William James went so far as to imply that for some "liberal" Christians of his day evolution understood as a progressive tendency is an *outgrowth* of religious sentiment which ends up replacing traditional religion (1958: 85 - 86).

[12] Cladistics is a school of systematics which groups organisms according to genetic lineage rather than appearance. Phenetics, on the other hand, is "[a]nother theory of classification, called phenetics --from a Greek word for appearance -- focuses on overall similarity alone and tries to escape the charge of subjectivity by insisting that phenetic classifications be based upon large suites of characters, all expressed numerically and processed by computer" (Gould 1983: 364). A cladogram is a diagram representing hereditary similarities. In general, the two species at the top right of a cladogram form a sister group. The next item is sister to those two (taken as a unit), the next to the previous three (again taken as a unit), and so on. What results is a "picture" which groups organisms according to their evolutionary proximity rather than on the basis of mere phenotypic similarities, which may be the results of convergence.

[13] "Order" is a handy notation for present purposes. At this point I do not mean to argue that the set of all propositions in evolutionary biology which involve the concept of fitness form a group in the mathematical sense. (Roughly, a group is a set plus an operation which, together, demonstrate closure, associativity, and the possibility of identity and inversion.) However, it will be argued that the set of "fitness propositions" has to be evaluated with something like an operation -- we'll call that something a function or algorithm -- in mind.

[14] There is no logical reason why a film could not repeat itself to some substantial degree. A 1992 film called "Groundhog Day," for instance, featured a character caught in what might be called an "infinite loop": every day he woke up at the same

time, to the same radio broadcast and the same weather. The people around him said and did the same thing unless he, himself, caused them to act differently. The protagonist sought a way to break out of the cycle, but there appeared to be no "exit condition," to a use a term from computer programming.

The point of mentioning this particular film is not to digress into a history of filmmaking. Rather, it is interesting to speculate that virtually any object which we experience in time could conceivably be so constituted as to be recursive -- that is, to repeat itself if analyzed on a sufficiently abstract level. This goes for movies, cosmologies (think of big bang-big crunch theories), religious accounts of the individual person's progress (Hindu notions of reincarnation or *samsara* as various organisms in what is essentially the same world; q.v. Matthews 126 - 128). The list could be expanded.

Where the sequentiality of a movie seems to differ from that which exists in a recursive object is that the essence of the movie itself is not inferable from any of its discrete moments. Of course we may take some images or phrases as representative of a particular film -- Clark Gable's "Frankly, Madam, I don't give a damn!" in "Gone with the Wind" as symbolizing the devil-may-care response of the ante-bellum South to its own weaknesses or Humphrey Bogart's "Here's looking at you, kid" as typifying a mature, duty-based love as opposed to an emotion grounded in selfish passion. And similarly with other creations of human genius: Hillel summarized the essence of Jewish scripture and culture by pointing to the so-called Golden Rule and asserting that "all the rest is commentary."

[15] Q.v. Hofstadter 1979: 68. The caption of the work is "Tiling of the plane using birds, by M. C. Escher (from a 1942 notebook).

[16] Hope renders the salient passage as "necessarily and always or even for the most part" (tr. 1961: 32).

[17]Putnam 1990, for instance, sketches his own position in this way (pp. 40 - 41):

> In my picture, objects are theory-dependent in the sense that theories with incompatible ontologies can both be right. Saying that they are both right is not saying that there are fields "out there" as entities with extensions and (in addition) fields in the sense of logical constructions. It is not saying that there are both absolute time points and points which are mere limits. It is saying that various representations, various languages, various theories, are equally good in certain contexts. In the tradition of James and Dewey, it is to say that devices which are functionally equivalent in the context of inquiry for which they are designed are equivalent in every way that we have a "handle on."

[18] Of course the theories of many economists are laced with what we can consider supervenient concepts. Ricardo comes naturally to mind because of the affinity between his and the Darwinian outlook. Erich Fromm notices this tie in his *Wege aus einer kranken Gesellschaft* (79 - 80):

> Hieraus geht eindeutig hervor, daß Freuds gesamte Sexualtheorie auf der anthropologischen Voraussetzung beruht, daß Rivalität und gegenseitige Feindseligkeit der menschlichen Natur innewohnend seien.
> Im Bereich der *Biologie* hat Darwin mit seiner Theorie vom wettstreitenden "Kampf ums Dasein" diesem Prinzip Ausdruck verliehen. Nationalökonomen wie Ricardo und die Manchester-Schule haben es in den Bereich der *Wirtschaft* übernommen.

[19] Clinical psychologist David Weeks wrestled with the concept of happiness in a similar fashion. Following a ten-year study of the phenomenon he calls eccentricity,

he was convinced that eccentrics tend to be happier than "normal" people. But that is a difficult claim to substantiate scientifically, since not just the definition of eccentricity but also that of happiness is so elusive. Weeks falls back on the same pattern of reasoning as the judge used in considering the definition of pornography (Weeks and James 1995: 39):

> We have asserted that it is not scientific to discuss something without defining it objectively. Nonetheless, any discussion of happiness, particularly when it is observed in other people, obliges us to fall back on the familiar formula that although we cannot prove its existence logically, or even say precisely what it is, we know it when we see it. Time and again, the eccentrics in our study clearly evinced that shining sense of positivism and buoyant self-confidence that comes from being comfortable in one's own skin. We [Weeks and the members of his research team] were always telling ourselves that if we could extract that happy essence and bottle it, we would be millionaires.

[20] Weber also cites a paper in print, Kim's 1995 "Emergence, Supervenience and Realization in the Philosophy of Mind," which I have not read.

[21] The authors do not introduce these examples to show that chance events can ruin theories where fitness is based on real numbers of offspring. Rather, the avowed purpose is to show that such definitions do not concur with the way biologists use the concept of fitness in doing real research. It is clear from Mills and Beatty's presentation, however, that there is no important difference between the two purposes, for the reason why scientists shy from the "real-offspring" definition of fitness is simply because such a definition would allow chance events to lead to counter-intuitive conclusions. See especially pp. 267 - 268 in Mills and Beatty 1979.

[22] Sober goes too far, however, when he claims that "nothing is left to chance" in artificial selection (1984a: 19). Animal breeders do *not* have complete control of their own breeding "experiments" because the wedding of genetic material from one parent with that of the other at fertilization is somewhat unpredictable. As Ridley succinctly puts it: "If all the genes on the chromosome were homozygous, the chromosomes after recombination would be identical to what they were before. But because the majority of genes are heterozygous the recombined chromosomes usually contain a different set of genes from the parental chromosomes. This is one of the main reasons why offspring differ from their parents. (Mendelian segregation is the other.)" (1985: 152). Contrary to Sober's statement, chance is inevitably a factor in artificial selection.

[23] Thanks to Prof. Carrier for making me aware of this important article.

[24] This of course assumes that we can infer a far-gone past from fossil evidence, a past which really existed. As Russell points out, there is no means of proving this logically. "Opponents of evolution, such as Edmund Gosse's father, urged a very similar argument against evolution. The world, they said, was created in 4004 B.C., complete with fossils, which were inserted to try our faith. The world was created suddenly, but was made such as it would have been if it had evolved. There is no logical impossibility in the view that the world was created five minutes ago, complete with memories and records. This may seem an improbable hypothesis, but it is not logically refutable" (1927: 7). Irrefutable or not, we won't entertain the elder Gosse's speculation here.

[25] The example:

> According to our insurance company, a white male American academic (type X) has a life expectancy of, say, 75 years. We say then that John, our colleague down the hall, has a life expectancy of 75 years if he instantiates this type. But, again according to our insurance company, an Irish Catholic polo player (type Y) has a life expectancy of only 69 years, and since John instantiates type Y as well, he has two incompatible life expectancies. Thus, it is clear that an organism which is an instance of a type that has a particular fitness value cannot be said to have that fitness value. [Ettinger *et al* (1990): 505]

[26] This is not to say that a thinker such as Dawkins would consider the choice of object arbitrary. Although at times Dawkins is compelled to speak of behavior or the individual as though it were the object of natural selection, it is always clear that he considers the gene to be the unit of selection, i.e., the thing which natural selection *qua* force moves. But elsewhere the author of this sentence gives primacy to the genetic origin of behavior as a trait: “Lewontin himself has expressed the point as well as anybody: ‘In order for a trait to evolve by natural selection it is necessary that there be genetic variation in the population for such a trait’ (Lewontin 1979)” (Dawkins 1982: 20). Moreover, it is clear that “trait” means not just a structural feature here, but can also refer to behavior, although perhaps behavior springs from (supervenes on) structure. “A gene ‘for’ behaviour X is a gene ‘for’ whatever morphological and physiological states tend to produce that behavior” (21).

[27] I mean that at some level all of these forces and phenomena are manifested in motion, not that they are themselves motion. It would be correct to say that many phenomena which do not seem to be instances of movement turn out to be just that, although the objects which move are not evident to the unaided senses. Forces such as gravity, on the other hand, differ from the motion which they can cause. Feynman makes a similar point this way: “...[I]n the early days [of the history of physics] there were phenomena of motion and phenomena of heat; there were phenomena of sound, of light, and of gravity. But it was soon discovered, after Sir Isaac Newton explained the laws of motion, that some of these apparently different things were aspects of the same thing. For example, the phenomena of sound could be completely understood as the motion of atoms in the air. So sound was no longer considered something in addition to motion. It was also discovered that heat phenomena are easily understandable from the laws of motion. In this way, great globs of physics were synthesized into a simplified theory. The theory of gravitation, on the other hand, was not understandable from the laws of motion, and even today it stands isolated from the other theories. Gravitation is, so far, not understandable in terms of other phenomena” (1985: 4).

[28] It’s interesting to compare this version of the problem with those appearing earlier in Copi’s text, particularly before the interpretative controversy appeared in Philosophy of Science. In 1968, the problem did not end with the question, “Are these two probabilities the same?” Instead, it ended with a parenthesized hint: “(These two probabilities are not the same!)” See Copi 1968: 433.

[29] Although Mills and Beatty’s article is almost always cited in treatments of fitness after 1979, there is one surprising exception which we will revisit below: Richard Dawkins’ *The Extended Phenotype* (1982). Dawkins devotes an entire chapter to distinguishing among five different types of fitness (chapter 10), but he neither mentions Mills and Beatty’s propensity interpretation explicitly nor includes their

article in his bibliography. He does, however, mention the issue of tendency or likelihood, in a way reminiscent of Mills and Beatty's thesis (Dawkins 1982: 184 - 5).

[30] At first glance the term might appear to be perfectly clear, and if it's not, why not just consult a basic reference work to find the consensus definition? Alas, most references are just as vague on the subject as Mills and Beatty are. Here's a sampling: "circular: ... 5. circuitous; roundabout; indirect: a circular argument (*Webster's Encyclopedic Unabridged Dictionary of the English Language*. 1994. New York: Gramercy Books: 268). "circular: ... 4: characterized by reasoning in a circle < ~ arguments>" (*Webster's New Collegiate Dictionary*. 1974. Springfield, Mass.: G. & C. Merriam Co.: 202). "circularity. Also circular reasoning or arguing in a circle. 1. Applied to ideas (arguments, reasoning, definitions) that repeat themselves. 2. Applied to arguments which assume the conclusion to be proven" (Angeles, Peter A. 1981. *Dictionary of Philosophy*. New York: Barnes & Noble: 39). This seems helpful, but after defining circular reason at greater length under the heading "definition, types of," Angeles notes that "most types of definition contain this circular quality" of repeating the *definiendum* in some form in the *definiens* (pp. 56 - 57). Presumably Mills and Beatty mean something very like this last definition, but their blithe treatment is more like the *Webster's* definitions, relying on the reader's intuitions rather than an explicit account. Since Sober unpacks the issue much more carefully than Mills and Beatty, we will treat the issue most fully when considering Sober's view.

[31] Mills and Beatty find forms of this argument in Williams and Ruse, but they rightly insist that the concept of fitness must nevertheless be clarified (1979: 266).

[32] Perhaps I misinterpret the authors' intention here, but it seems to me that they see the definition of fitness as the linchpin of the controversy about the integrity of evolution theory as a whole. "But the fact is that there is a major problem in the foundations of evolutionary theory which remains unsolved, and which continues to give life to the debate. The definition of fitness remains in dispute, and the role of appeals to fitness in biologists' explanations is a mystery" (1979: 264; in Sober 1984: 37).

[33] "Fitness may be predicated of individual organisms, and (in a somewhat different sense) of phenotypes and genotypes" (1979: 267; 39 in Sober 1984c).

[34] A charming example is offered by entomologist May Berenbaum (1995: 173) and even repeated by Albrecht (1997) in Germany's bestselling weekly newspaper. I offer Albrecht's summary to demonstrate the meme-like quality of this observation:

> Nicht jedermann ist mit der Tatsache vertraut, daß der Darmtrakt einer einzelnen Termite pro Tag zwischen 0,24 und 0,59 Mikrogramm Methan produziert. 200 Billiarden dieser unermüdlichen Holzfresser bevölkern den Erdball, und daraus kann man ableiten, daß furzende Termiten bis zu dreißig Prozent des gesamten Methangehalts der Atmosphäre erzeugen -- ein beachtlicher Beitrag zum Klimageschehen. (1997: 33)

[35] A similar distinction is that between *act* and *rule* utilitarianism, in which the greater good is sought, respectively, by observing what would be the best outcome moment by moment, act by act, without paying attention to longer-term consequences (act utilitarianism), or by following a course of action whose outcome appears not to be the best from an immediate perspective but which will lead to the best outcome in the longer term. An act utilitarian might attempt to realize (in Hutcheson's phrase) "the greatest good for the greatest number" by the vigilante murder of a known Mafia

kingpin whose misdeeds have not been proved in a court of law. An act utilitarian might forego the opportunity to exact such frontier justice, thus allowing the kingpin to continue to damage the society, on the grounds that guaranteeing due process to all citizens will lead to a better society in the long term.

[36] The substantial point here should not be affected by the fact that the two examples in this section -- Cinderella and the rabbit -- are both hypothetical. I would not wish to be accused of turning a bear into a whale, a phenomenon which Darwin sketched in a hypothetical example which he came to regret having ever published (Gould 1995: 359 - 360).

[37] Knipe does not share this view. He attempts to identify the defining characteristics of Hinduism. However, it becomes clear in the course of his presentation that to isolate the common features of the various outlooks subsumed under the heading, he must work at a very abstract level. One wonders whether practitioner of various "Hindu" faiths would recognize their belief system in Knipe's schematism of Hinduism.

[38] While Sober's "fog" obscures truth in general, the fog which settles over Maine in King's "The Mist" claims, as its first certain victim, a character named (appropriately in this context) Norm.

[39] The theory of evolution is intended to be descriptive, of course, not normative. Some have tried to bridge the gap between science and normative ethics (E. O. Wilson has been widely cited -- and misinterpreted -- in this arena), but this sort of speculation should not be seen as part of the theory of evolution *per se*. Nonetheless, one can see certain abstract parallels in the patterns of thought central to Darwinian evolution and utilitarianism. Key to both theories is a notion of advantage which starts at the individual level but can be extrapolated to apply to the collective. If something is of advantage (defined as leading to the happiness of the individual) in an ethical sense, then it should be perpetuated if there are no "side effects"-- so one possible reading of Mill's utilitarianism. Darwinian evolution hypothesizes that if some feature was of advantage, then it was perpetuated assuming there were no counteractive forces operative.

[40] That this is Wachbroit's intention becomes clear in the course of his article. For instance, he asserts that

> ...Nagel claims that the heart is necessary for circulatory blood if we restrict our attention only to *normal* organisms (1977, 292). Although an organism can have circulating blood without having a heart, this would not be a normal organism.
>
> Clearly, Nagel's proposal employs the biological concept of normality. Statistical or evaluative conceptions of normality would plainly not be plausible candidates for Nagel's strategy [583].

Contrary to Wachbroit's assertion here, there is a good case to be made that Nagel's "strategy" is very much statistical: Within the range of organisms which Nagel considers, *most* which have circulating blood also have hearts. Therefore, Nagel claims that having a heart is normal for organisms with circulating blood.

[41] The adjective "homologous" is applied in two distinct senses. It can be applied to phenotypical traits such as bone structures in the limbs of various reptiles and mammals, which is the sense used here. But chromosomes, too, are called homologous if they bear the same genes (Futuyma 1986, p. 552).

[42] I have tried in vain to find a printed source documenting this account of Duvalier's conversation with journalists. The story was told by the Berlin-based journalist Hans Christoph Buch in a talk entitled "Die neue Weltordnung: Ethnizität und Demokratie in Liberia, Ruanda, Bosnien und anderswo," given at the Amerika Haus in Frankfurt, Germany, September 28, 1995. Apparently Buch was not present when Duvalier made the alleged remarks but had the story at second-hand as well.

[43] Some authors (e.g., Brandon 1990) use the term "adaptation" as a synonym for "fitness." In itself there is nothing wrong with stipulating that two terms are to be used as synonyms, but in the case of "fitness," "adaptation," and their derivatives, a possible source of confusion arises. One can speak of adaptation in general, or of a specific adaptation, meaning a change in morphology or behavior. An organism can also be called adapted with respect to a certain environment. Fitness, on the other hand, is more general in its usage. While one can speak of fitness and adaptation as general properties, there seems to be no one-word derivative of *fitness* which corresponds to the word *adaptation* meant as a specific change in morphology or behavior. For instance, one might say that leaflessness in plants inhabiting a xeric environment is "an adaptation" (Futuyma 1986, Ch. 9), but one cannot call the same phenomenon "a fitness." It seems clear that the terms adaptation and fitness are in fact not perfect synonyms.

[44] Cf. Cracraft 1989: 31: "Species definitions are used much like evangelicals use the Bible: as a putative guide to reality. Species *are* what the definition says they are; the entity before us satisfies the definition, therefore it must *be* a species."

[45] In a television interview, Johnson said: "Now, Steve Gould, the Harvard professor who is the best known popularizer in America, wrote a paper for a professional audience in 1981, in which he said, 'Neo-Darwinism is effectively dead as a general theory, despite its persistence as textbook orthodoxy.'

"...And then he [Gould] said, 'Well, things change suddenly. You don't see the evidence of it because it happens in an instant of geological time,' and he hinted very strongly, without really committing himself, to what are called the Saltationist views that have been associated with some earlier heretics [i.e., evolutionists who dissent from the "party line" of *gradual* evolution], in which there may be an evolution by big jumps...

"Well, you see, theories like punctuated equilibrium have come up really since the beginning of Darwinism. T.H. Huxley was much along the same line. They come up whenever a Darwinian is thinking primarily about the fossil record. They always come from the fossil people. And it's because the fossil record doesn't show any Darwinian transitions, you see. So they're trying to find a way to reconcile the theory with that record. Now as long as they're doing that within the family, it's all right. But the outside critic, who's not polite, comes in and says, 'Well, this means that you're talking about magical kind of jumps. You don't really have a materialistic way to explain the development of eyes and lungs and brains and all those tremendously intricate things that look like they needed a designer.' So at that point everybody runs back to orthodox Darwinism because that's the only way you can get what they call the blind watchmaker" ("Firing Line" 1991, pp. 5-6).

[46] Gould offers further examples to make his point. "In 1966, Jared Diamond published a more extensive study of the Fore people of New Guinea. They have names for all the Linnaean bird species in their area. Moreover, when Diamond brought seven Fore men into a new area populated by birds they had never seen, and

asked them to give the closest Fore equivalent for each new bird, they placed 91 of 103 species into the Fore group closest to the new species in our Western Linnaean classification" (1980: 108). Anthropologist Ralph Bulmer "could only find four cases (2 percent) of inconsistency in the [New Guinean] Kalam catalog of 174 vertebrate species, spawning mammals, birds, reptiles, frogs and fishes" (1980: 209). Anthropologist Brent Berlin's, and botanists Dennis Breedlove and Peter Raven's "complete catalog [of species recognized by the Tzeltal Indians of Chiapas, Mexico] contains 471 Tzeltal names. Of these, 281, or 61 percent, stand in one-to-one correspondence with Linnaean names. All but 17 of the rest are, in the authors' terms, 'undifferentiated' -- that is, the Tzeltal names refer to more than one Linnaean species. But, in more than two-thirds of these cases, the Tzeltal use a subsidiary system of naming to make distinctions within the primary groups, and all these subsidiaries correspond with Linnaean species. Only 17 names, or 2.6 percent, are 'overdifferentiated' by referring to part of a Linnaean species. Seven Linnaean species have two Tzeltal names, and only one has three -- the bottle gourd *Lagenaria siceraria*. The Tzeltal distinguish bottle gourd plants by the utility of their fruits ... (1980: 210).

[47] Gould adds in parentheses a note that at first blush seems contradictory to his stated doubt about the dependence of the perception of species on human neurology: "I do not, of course, deny that our propensity for classifying in the first place reflects something about our brains, their inherited capacities, and the limited ways in which complexity may be ordered and made sensible. I merely doubt that such a definite procedure as classification into Linnaean species could reflect the constraints of our mind alone, and not of nature" (1980: 212). The key term here is "propensity." Gould allows that something about our human-ness motivates us to classify; but he will not allow that the details of Linnaean classification are imposed by a quirky humanity on a nature which is essentially other than how we perceive it.

What we're dealing with here is a fairly simple distinction between what really is versus what is imposed in the act of perception (ignoring for a moment finer epistemological questions of whether and how we can separate the two realms). But it is worth considering whether a subdistinction should be drawn between locally versus remotely imposed taxonomies. It seems fairly clear that a key concept here is that of interest: if an observer has a particular interest (purpose), then her perception of reality will be affected by that interest. She may rightly be said to carve nature at the joints, but she'll select which joints to carve and reject others. There is enough evidence that this is a general rule in human endeavor so that an *a fortiori* argument for the existence of this subdistinction can be made. Historically, for instance, the geography of large-scale nationalism differs from that of early settlement. "The boundaries of most colonial-era land grants had been haphazard, reflecting colonial America's uneasy combination of speculation and the need for lasting communities. Along the Atlantic coast there are still county and township lines as twisting as any that Europe has inherited from its medieval past. West of the Appalachian chain, however, the lines become straight, oriented directly toward the North Pole. Literally, they are imposed from above for the sake of quick sale, clarity of possession, and easy transfer" (Countryman 1996: 77).

[48] It is interesting that during his expedition across the continent (1803 - 1806), his evolutionary musings were about language rather than species. This, despite the fact that his contributions as a naturalist were extraordinary. "He had discovered and

described 178 new plants...and 122 species and subspecies of animals" (Ambrose 1996: 394). Ambrose laments the fact that Lewis did not publish his journals, saying that he "had cheated himself out of a rank not far below Darwin as a naturalist" (1996: 470).

[49] Mayr continues his line of thought in the next paragraph: "No one resented Darwin's independence of thought more than the philosophers....Darwin had violated all the rules of the game by placing his argument entirely outside the traditional framework of classical philosophical concepts and terminologies....No other work advertised to the world the emancipation of science from philosophy as blatantly as did Darwin's *Origin*. For this he has not been forgive to this day by some traditional schools of philosophy. To them, Darwin is still incomprehensible, 'unphilosophical,' and a bête noire" (Mayr 1964: xi - xii). Mayr has his own, doubtless legitimate rhetorical reasons for using dramatic phrases such as "completely eliminating the last remnants of Platonism" and "placing his argument entirely outside the traditional framework." For present purposes we need only bear in mind that Darwinian evolution requires a taxonomy grounded in a regularity which can be characterized in Platonic terms.

[50] This is not to suggest that Kuhn was the first to offer the thesis that scientific theories are theory-laden. In 1927 Englishman Russell offered a tongue-in-cheek version of theory-ladenness in accounts of animal behavior in which national culture infects observation: "One may say broadly that all the animals that have been carefully observed have behaved so as to confirm the philosophy in which the observer believed before his observations began. Nay, more, they have all displayed the national characteristics of the observer. Animals studied by Americans rush about frantically, with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think, and at last evolve the solution out of their inner consciousness" (1927: 33).

[51] Ford, E. B. $^{3}$1971. *Ecological Genetics*. London: Chapman and Hall.

[52] Of course many principles useful to the reconstruction of biological history can also be used to speak of what will happen under idealized conditions. For instance, the competitive exclusion principle (also called Gause's axiom) asserts that species cannot coexist indefinitely if they use the same resources. This principle may be used to explain why a species which was apparently well adapted to many of the features of its environment became extinct, but the principle may also be used to predict what will happen in the future (Futuyma 1986, esp. p. 30).

[53] Speaking of the earliest life forms on the planet, Gaylord Simpson writes: "Indeed it is improbable that the discovery of [any of their fossilized] remains, if any do exist, would greatly advance knowledge of how life originated. At this lowest level little could be learned from the preserved form: the problem is physiological, not morphological, and it seems that form must develop above the molecular level before it can serve as a particularly useful clue to function" (1967, p. 15).

[54] Durrant, A. 1962. "The Environmental Induction of Heritable Change in *Linum*." *Heredity* 17: 27 - 61. Cullis, C. A. 1983. "Environmentally Induced DNA Changes in Plants." *CRC Critical Reviews in Plant Science* 1: 117 - 131.

[55] "Denn Evolution ist ein einzigartiger, prinzipiell nicht wiederholbarer Prozeß, der sich der direkten Analyse entzieht und nur Rekonstruktionen erlaubt. Damit steht der Biologe übrigens nicht alleine da: Alle historischen Disziplinen, wie Kosmologie,

Archäologie oder Mittelalterliche Heraldik, haben dasselbe Problem" (Haszprunar 1994: 131).

[56] The linguist at center stage in this paper, Roger Lass, has already pointed out the relationship between taxonomy and historical linguistics:

> "We also share a concern with biologists if we're interested in taxonomy, which is an integral part of historical linguistics. A question like 'Is Gothic "really" North Germanic?' is very like 'Is the Livingstone Daisy (Bokbaai vygie) "really" a *Dorotheanthus*, or should it stay a *Mesembryanthemum*?' Both questions involve -- among other things -- weighting of phenotypic characters. In a linguistic context, how much should (say) Holtzmann's Law count in assigning a Germanic language to a particular subfamily? Such questions play the same theoretical and specifically historical roles; since, if we assume the aims of taxonomy to be phylogenetic, claims about ultimate 'genotypes' are involved in both cases" (Lass 1986: 35, n. 1).

But so far as I know Lass never wielded the notion of taxonomy as a means of analyzing the GVS in particular. Had he done so in his 1992 paper, it is hard to imagine that he would have abandoned the final and outer efficient causality evident in his earlier work (e.g., 1976) without a good deal more justification.

[57] Throughout this chapter "traditional version" means any theory which conceives of English long vowel shifts between a period beginning roughly with ME and ending with EModE as constituting a single phenomenon. (The period is difficult to specify precisely, since some authors date the beginnings of what is traditionally called the GVS in OE -- e.g., Martinet 1955, Stockwell and Minkova 1988a -- and it may be true that long vowels continue to shift in an ongoing way, i.e., one connected with the earlier changes.

[58] In this paper we concentrate on efficient and final causality, thus it may be helpful to remind ourselves of Aristotle's introduction to these two types of cause:

> In another [[sense, "a cause"]] means (3) that from which change or coming to rest first begins; for example, the adviser is a cause, and the father is the cause of the baby, and, in general, that which acts is a cause of that which is acted upon, and that which brings about a change is a cause of that which is being changed.
>
> Finally, it means (4) the end, and this is the final cause [that for the sake of which]; for example, walking is for the sake of health. Why does he walk? We answer, "In order to be healthy"; and having spoken thus, we think that we have given the cause. And those things which, after that which started the motion, lie between the beginning and the end, such as reducing weight or purging or drugs or instruments in the case of health, all of them are for the sake of the end; and they differ in this, that some of them are operations while others are instruments. (Tr. Apostle 1980: 29-30; my double square brackets, his single square brackets)

Another translation suggests that the efficient cause is the "prime" agent effecting a change:

> Then again (3), there must be something to initiate the process of the change or its cessation when the process is completed, such as the act of a voluntary agent (of the smith, for instance), or the father who begets a child; or more generally the prime, conscious or unconscious, agent that produces the effect and starts the material on its way to the product, changing it from what it was to what it is to be. (Wicksteed and Cornford 1980: 129 - 131).

A point made in the text of this paper is highlighted by the translations' difference in this respect: How do we decide which of the many identifiable agents of a given change is the "prime" motive force? It seems that we cannot.

[59] Since my MS Windows synthesis of ASCII and ANSI cannot reproduce the entire IPA alphabet, I use ⊃ to represent the rounded half-open back vowel and ∂ for the half-close central vowel.

[60] The "simplicity" and "elegance" which Stockwell and Minkova cite can be read as aesthetic criteria, but it is just as possible to read into these words a pragmatic appeal -- a usefulness in organizing data that would otherwise be messier and therefore harder to grasp. Einstein seemed to have thought along these lines as well:

> "...concepts which have proved useful for ordering things assume so great an authority over us, that we forget their terrestrial origin and accept them as unalterable facts. They then become labeled as 'conceptual necessities,' '*a priori* situations,' etc. The road to scientific progress is frequently blocked for long periods by such errors. It is therefore not just an idle game to exercise our ability to analyse familiar concepts, and to demonstrate the conditions on which their justification and usefulness depend, and the way in which these developed, little by little, from the data of experience. In this way they are deprived of their excessive authority." (Quoted in Sober 1988: ix)

[61] Martinet, André 1955. *Economie des changements phonétiques*. Bern: Francke.

[62] Although it seems to me axiomatic that resemblance is a composite of factors imposed by the observer and of qualities inherent in what is observed, I am aware that the position is problematic and that some respected thinkers seem to lean to what I have called the strong version of Lass's Constellation Theory (e.g., Goodman 1976: 34 - 39).

[63] Such a social cause is sketched by Leith, for example:

> "We have now outlined most of the changes involved in the shift. But we have yet to suggest the mechanism. As we have seen, a relatively high variant of the vowel in *mate* was associated with the speech of Essex and Kent, and as we saw in chapter two, Kentish was a stigmatized dialect. The London bourgeoisie, then, would want to distance its own pronunciation from that of the lower class, which was constantly being swelled by immigrants from these areas. One way of doing this was to raise the vowel of *mate* even higher than that of the lower-class variant; and raising of the lowest vowel in the system would necessitate raising all the vowels above and, ultimately, pushing the vowel of *tide* into a diphthong. It seems that in the speech of the bourgeoisie, the vowel of *mate* was raised to a height close to that of *meat*, so that some observers actually recorded a merger of the two sounds...The aristocracy,

> now no longer able to distance itself with the use of French, seems at first to have kept *mate*, *meat*, and *meet* distinct. At the other extreme, a third system had merged *meat* and *meet*. It appears that this was the lower-class pattern: and the fact that it is this that eventually formed the pattern for the future prestige accent need not surprise us. As is often the case, the unacceptable yesterday becomes the acceptable today." [Leith 1983: 149]

[64] Gould, in turn, borrowed his title (1983) -- "What, if Anything, is a Zebra?" -- from Albert E. Wood's 1957 article "What, if Anything, is a Rabbit?"

[65] In 1977 Gould and Niles Eldredge published their essay "Punctuated equilibria: the tempo and mode of evolution reconsidered." (*Paleobiology* 3: 115 - 151). Although viewed as building on earlier theories, the essay is considered the primary expression of the theory by both supporters and detractors. Gould remains the outstanding example in the first category, still energetically defending himself and his insights, most recently in 1995: 123 - 132. Recent detractors are Dennett 1995: esp. 282 - 312 and Ruse 1995: 70 - 105. Ruse attempts (unsuccessfully, in my view) to show that Gould and his theory are less influential than commonly thought.

[66] Dennett 1995 and Ruse 1995 are especially acerbic opponents (n. 10 above); Futuyma 1986: 401 - 409 summarizes various views and suggests that Darwinian gradualism and some versions of punctuated equilibrium can be seen as consistent.

[67] My translation: "...die in der Sprache und ihrer Entwicklung die entscheidende Rolle spielen..."

[68] Boisson, Claude 1982. "Remarques sur la chronologie interne du grand changement vocalique en anglais", *Apports français à la linguistique anglaise* (*Travaux* 35, CIEREC, Université de Saint-Etienne).

[69] Lass does not say this explicitly, but he seems to sense that our perception of sound changes such as those occurring in the GVS are unfairly manipulated by a sort of binary fever. This sickness exists when we insist that formal mappings be binary. Thus Lass offers us a hypothetical language which is affected by two rules:

$$\text{(a)} \quad \begin{matrix} \text{e} \\ \\ \text{o} \end{matrix} \quad \rightarrow \quad \begin{matrix} \varepsilon \\ \\ \supset \end{matrix}$$

$$\text{(b)} \quad \begin{matrix} \text{i} \\ \\ \text{u} \end{matrix} \quad \rightarrow \quad \begin{matrix} \text{e} \\ \\ \text{o} \end{matrix} \qquad \text{(Lass 1976: 69)}$$

Lass then restates these rules in binary fashion (binary because a single rule deals with at most two heights at a time).

$$\text{(a)} \quad \begin{matrix} \text{V} \\ +\text{high} \\ +\text{mid} \end{matrix} \quad \rightarrow \quad [-\text{high}]$$

$$\text{(b)} \quad \begin{matrix} \text{V} \\ +\text{high} \\ -\text{mid} \end{matrix} \quad \rightarrow \quad [+\text{mid}]$$

(ibid.)

(The notation is Wang's (1968).)

"But there is clearly a generalization here, which is that 'nonlow vowels lower one height'. I maintain that it is perverse to deny the existence of processes like 'lowering by one height', and that accordingly there should be rules of the form:"

$$\begin{matrix} \text{V} \\ \\ +\text{height}^{n} \end{matrix} \quad \rightarrow \quad [+\text{height}^{n-1}]$$

(ibid.: 70)

By limiting ourselves to binary rules, we tend to "undergeneralize" (or so claims Lass), which in turn means that we tend to view sound changes as discrete rather than as parts of unified wholes.

In Lass's opinion, metarules are also necessary to represent real aspects of the GVS which binary rules capture, at best, inadequately. For instance, he offers the following metarule and associated caveats to show how raising and diphthongization relate to one another:

a. *Metarule*: $\text{V}_1 \quad \text{V}_2 \quad \rightarrow \quad [+\text{raise}]_1 \; [+\text{raise}]_2$

b. Condition: No collapse if (a) [with some exceptions]

c. Implication: If ~ (b), then:

$$\begin{matrix} \text{V} \\ \\ +\text{high} \\ 1 \end{matrix} \qquad \begin{matrix} \text{V} \\ \\ +\text{high} \\ 2 \end{matrix} \quad \rightarrow \quad \begin{matrix} [-\text{high}] \\ \\ 1 \qquad 2 \end{matrix}$$

(ibid.: 71)

One caution: Lass's metarules are not to be confused with Chomsky and Halle's formalizations in respect to basic type and function. "I am not saying that metarules...are ACQUIRED by children; they are EVENTS in the history of a language that precipitate system-wide changes. The output that children use for acquisition, on the other hand is of course directly generated by the rules in their parents' grammars that implement metarules" (Lass 1976: 72). "...[M]etarules are not, in the usual way, 'added to' grammars. They are EVENTS in linguistic history whose nature is induced from their effects..." (ibid. 84).

[70] Defining schools of thought is always a controversial undertaking. There's no need to attempt a detailed exposition here, but in case there be any question as to what definitions I implicitly accept, I would point to simple categorizations of positions on scientific theories (presumably applicable to phonology as well as other physical sciences). Laudan (1990), for instance, describes positivist, pragmatist, realist, relativist schools. In Laudan's dialogue, the spokesmen for the positions are, respectively, Rudy Reichfeigl (Rudolf Carnap-Hans Reichenbach-Herbert Feigl?), Percy Lauwey (C.S. Peirce-Larry Laudan?), Karl Selnam (Karl Popper-Wilfrid Sellars-Hillary Putnam) and Quincy Rortabender (Willard v. O. Quine-Richard Rorty-Paul Feyerabend?).

[71] One of my teachers bemoaned the devastating effect President Jimmy Carter, a native of Georgia, had had on the nationwide pronunciation of "nuclear." Following a single, nationally televised speech, more than half of her students -- or so the teacher claimed -- adopted Carter's pronunciation. I thought the teacher was exaggerating until I read Bryson's 1990 mention of a one-time Carterism (confusing "flaunt" and "flout"). Considering the attention this single instance received, it is conceivable that a president's continued non-standard pronunciation of a word over four years might indeed have a dramatic effect among speakers of American English.

# Bibliography

Aiello, Leslie and Christopher Dean. 1990. *An Introduction to Human Evolutionary Anatomy*. London: Academic Press.

Aiken, Henry D. (ed) 1948. *Hume's Moral and Political Philosophy*. New York: Hafner Press.

Albrecht, Jörg. 1997. "Königin der Käfer." *Die Zeit*. 21.März 1997: 33.

Alioto, Anthony M. 1987. *A History of Western Science*. Englewood Cliffs, NJ: Prentice-Hall.

Allen, T. O., N. T. Adler, J. H. Greenberg, and M. Reivich. 1981. "Vaginocervical Stimulation Selectively Increases Metabolic Activity in the Rat Brain." *Science* 211: 1070 - 1072.

Alston, William P. (1963). "Meaning and Use." Reprinted in Feigl *et al* (1972): 243 - 256.

Ambrose, Stephen E. 1994. *D-Day, June 6, 1944. The Climactic Battle of World War II.* New York: Simon and Schuster.

---. 1996. *Undaunted Courage. Meriwhether Lewis, Thomas Jefferson and the Opening of the American West.* New York: Simon and Schuster.

Angeles, Peter A. 1981. *Dictionary of Philosophy*. New York: Barnes and Noble.

Árdal, Páll S. 1989. *Passion and Value in Hume's Treatise*. Edinburgh: Edinburgh University Press.

Aristotle. (1961). *Aristotle's Physics*. Translated by Richard Hope. Lincoln, Neb. and London: University of Nebraska Press.

--- (1980). *Aristotle's Physics*. Translated by Hippocrates G. Apostle. Grinnell, Iowa: The Peripatetic Press.

--- (1980). *The Phyiscs* Books I - IV (Vol. 1). Translated by Philip H. Wicksteed and Francis M. Cornford. (The Loeb Classical Library) Cambridge, Massachusetts: Harvard University Press.

--- (1934). *Nicomachean Ethics*. Translated by H. Rackham. (The Loeb Classical Library) Cambridge, Massachusetts: Harvard University Press.

Axelrod R. and W. D. Hamilton. 1981. "The Evolution of Cooperation." *Science* 211: 1290-1296.

Ayala, Francisco J. 1978. "Mechanisms of Evolution." *Scientific American* (Sep): 56 - 69.

Baldwin, J. M. 1896. "A New Factor in Evolution." *American Naturalist*, 30: 441-51, 536-53.

Balme, D. M. 1975. "Aristotle's Use of Differentiae in Zoology." In Barnes et al (eds) 1975. First published in S. Mansion (ed) *Aristote et les problèmes de méthode*. Louvain: 1961, 195 - 212.

Barnes, Jonathan. 1987. *Early Greek Philosophy*. London: Penguin.

---, Malcolm Schofield, and Richard Sorabji. 1975. *Articles on Aristotle 1. Science*. London: Gerald Duckworth.

Barrow, John D. and Frank J. Tipler. 1986. *The Anthropic Cosmological Principle*. Oxford: Oxford University Press.

Baugh, Albert C. and Thomas Cable. 1993. *A History of the English Language* (4th revised edition). London: Routledge.

Beatty, John 1980. "Optimal-design Models and the Strategy of Model Building in Evolutionary Biology." *Philosophy of Science* 47: 532 - 561.

---. 1984. "Chance and Natural Selection." *Philosophy of Science* 51: 183 - 211.

--- and Susan Finsen. 1989. "Rethinking the Propensity Interpretation: A Peek Inside Pandora's Box." In Ruse (ed.) 1989: 17 - 30.

Benditt, John. 1988. "Cousins or Brothers?" *Scientific American* 258: 18.

Benenson, F.C. 1984. *Probabilty, Objectivity and Evidence*. London: Routledge and Kegan Paul.

Bentley, Jon. 1988. *More Programming Pearls. Confessions of a Coder*. Reading, Mass.: Addison-Wesley.

Berenbaum, May. 1995. *Bugs in the System: Insects and Their Impact on Human Affairs*. Reading, Mass.: Addison - Wesley.

Bergamini, David. 1980. *Mathematics* (rev. ed.) Alexandria, VA: Time-Life Books.

Bindra, Dalbir and Francine G. Patterson, H.S. Terrace, L.A. Petitto, R.J. Sanders, T.G. Bever. 1981. "Ape Language." *Science* 211: 86 - 88.

Birdwhistell, Ray L. 1952. *An Introduction to Kinesics*. Louisville, KY: University of Louisville Press.

Bloomfield, Morton W. and Leonard Newmark. 1963. *A Linguistic Introduction to the History of English*. Westport, Conn.: Greenwood.

Bowlby, John. 1990. *Charles Darwin*. New York: Norton.

Bowler, Peter. 1989. *Evolution*: *The History of an Idea* (Rev. ed.). Berkeley and Los Angeles: University of California Press.

Boyer, Paul. 1992. *When Time Shall Be No More. Prophecy Belief in American Culture*. Cambridge, Mass.: Belknap Press.

Bramly, Serge. 1991. *Leonardo. Discovering the Life of Leonardo da Vinci.* Translated by Siân Reynolds. New York: HarperCollins.

Brandon, Robert N. 1978. "Adaptation and Evolutionary Theory." *Studies in the History and Philosophy of Science*, 9, no. 3, pp. 181 - 206. Reprinted in Sober (1984b), pp. 58 - 82.

Brandon, Robert N. 1990. *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Brandon, Robert N. and John Beatty. 1984. "Discussion: The Propensity Interpretation of 'Fitness' -- No Interpretation is no Substitute." *Philosophy of Science* 51: 342 - 347.

Brandon, Robert N. and Scott Carson. 1996. "The Indeterministic Character of Evolutionary Theory: No 'No Hidden Variables Proof' but No Room for Determinism Either." *Philosophy of Science* 63: 315 - 337.

Bryson, Bill. 1990. *The Mother Tongue*. New York: William Morrow and Co.

Burfoot, Amby. 1992. "White Men Can't Run." *Runner's World* 8: 89 - 95.

Burkhardt, Frederick. 1996. *Charles Darwin's Letters. A Selection 1825 - 1859.* Cambridge: Cambridge University Press.

Calvin, William H. 1994. "The Emergence of Intelligence." *Scientific American* (Oct): 100 - 107.

Campbell, J. H. 1982. "Autonomy in Evolution." In R. Milkman, ed., *Perspectives on Evolution*, pp. 190-201. Sunderland, Mass.: Sinauer Associates.

Caplan, Arthur L. 1977. "The Nature of Darwinian Evolution: Is Dawinian Evolutionary Theory Scientific?" In Godfrey, L., ed. *What Darwin Began.* Boston: Allyn and Bacon.

Carrier, Martin. 1992. "Cavendishs Version der Phlogistonchemie oder: Über den empirischen Erfolg unzutreffender theoretischer Ansätze." In Jürgen Mittelstraß and Günter Stock (Hrsgs) 1992. *Chemie und Geisteswissenschaften. Versuch einer Annäherung.* Berlin: Akademie Verlag: 35 - 52.

---. 1994. *The Completeness of Scientific Theories. On the Derivation of Empirical Indicators within a Theoretical Framework: The Case of Physical Geometry.* (The University of Western Ontario Series in the Philosophy of Science: v. 53) Dordrecht: Kluwer Academic.

Cavenee, Webster K. and Raymond L. White. 1995. "The Genetic Basis of Cancer." *Scientific American* (Mar): 72 - 81.

Chalmers, David J. 1995. "The Puzzle of Conscious Experience." *Scientific American,* 273: 62 - 68.

Charig, Alan J. and Angela C. Milner. 1986. "Baryonyx, a remarkable new theropod dinosaur." *Nature* 324 (27 Nov): 359 - 361.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper and Row.

Christianson, Gale E. 1984. *In the Presence of the Creator. Isaac Newton and his Times.* New York and London: Free Press.

Christie, Agatha. (1939.) *And Then There Were None.* New York: Berkley, 1991.

Christie, William M., Jr. 1982. "Synchronic, Diachronic, and Panchronic Linguistics." In Maher *et al* (eds) 1982: 1 -10.

Clark, Ronald W. 1984. *The Survival of Charles Darwin. A Biography of a Man and an Idea.* New York: Random House.

Clausewitz, Karl von. 1960. *The Principles of War.* (Tr. Hans W. Gatschke). Harrisburg, Penn.: Stackpole.

Cooper, James Fenimore. 1986. *The Last of the Mohicans.* New York: Norton.

Coppleston, Frederick 1953. *A History of Philosophy* (vol 1). New York: Doubleday.

Cohen, I. Bernard. 1995. *Science and the Founding Fathers. Science in the Political Thought of Jefferson, Franklin, Adams and Madison.* New York and London: Norton.

Cohen, Stanley N. and James A. Shapiro. 1980. "Transposable Genetic Elements." *Scientific American* 242: 40 - 49.

Conte, S. D. and Carl de Boor. 1980. *Elementary Numerical Analysis. An Algorithmic Approach* (3rd ed). (International Series in Pure and Applied Mathematics) New York: McGraw-Hill.

Cook, Martin L. 1991. *The Open Circle. Confessional Method in Theology.* Minneapolis, Minn.: Fortress Press.

Copi, Irving M. 1968. *Introduction to Logic (3rd ed.).* New York: Macmillan.

Copi, Irving M., and Carl Cohen. 1990. *Introduction to Logic (8th ed.*). New York: Macmillan.

Countryman, Edward. 1996. *Americans. A Collision of Histories.* New York: Hill and Wang.

Cowgill, Ursula M. 1970. "The People of York: 1538 - 1812." *Scientific American* (Jan): 104 - 112.

Cracraft, Joel. 1989. "Species as Entities of Biological Theory." In Ruse (ed.) 1989: 31 - 52.

Crews, David. 1994. "Animal Sexuality." *Scientific American* (Jan): 108 - 114.

Crichton, Michael and David Koepp. 1993. "Jurassic Park" (screenplay). Universal City, Calif.: Universal Studios.

Crosby, Alfred. 1996. *The Measure of Reality. Quantification and Western Society, 1250 - 1600.* Cambridge: Cambridge University Press.

Dale, A.I. 1974. "On a Problem in Conditional Probability." *Philosophy of Science*, 41: 204 - 206.

Darwin, Charles. 1859. *On the Origin of Species.* Cambridge, Mass.: Harvard University Press, 1964.

Davis, Morton D. 1970. *Game Theory.* New York and London: Basic Books.

Davis, Philip J. and Reuben Hersh. 1986. *Descartes' Dream. The World according to Mathematics.* San Diego: Harcourt Brace Jovanovich.

Dawkins, Richard. (1976.) *The Selfish Gene.* Oxford: Oxford University Press, $^{2}$1989.

---. 1982. *The Extended Phenotype*. Oxford and San Francisco: W.H. Freeman.

---. 1986. *The Blind Watchmaker*. Essex: Longman Scientific & Technical.

---. 1995. *River Out of Eden*. London: Phoenix.

Dennett, Daniel C. 1991. *Consciousness Explained.* London: Penguin.

---. 1995. *Darwin's Dangerous Idea. Evolution and the Meanings of Life*. New York: Simon and Schuster.

Descartes, René. (1637). *La Geometrie*. In David Eugene Smith and Marcia L. Latham (trs). 1954. *The Geometry of René Descartes* (with a facsimile of the first edition). New York: Dover.

---. (1641). *Meditations on First Philosophy*. In *The Essential Works of Descartes* (Tr. Lowell Bair). New York: Bantam, 1961.

Desmond, Adrian and James Moore 1991. *Darwin*. New York: Warner Books.

Deutsch, David and Michael Lockwood. 1994. "The Quantum Physics of Time Travel." *Scientific American* (Mar): 68 - 74.

Dewey, John. 1910. *The Influence of Darwin on Philosophy and Other Essays in Contemporary Thought*. Bloomington: Indiana University Press.

---. 1922. *Human Nature and Conduct. An Introduction to Social Psychology*. New York: Modern Library.

Diamond, Jared M. 1987. "Aristotle's theory of mammalian teat number is confirmed." *Nature* 325 (15 Jan): 200.

---. 1996. *Germs and Steel. Investigations in the Science of Human History*. New York: Norton.

Dickerson, Richard E. 1978. "Chemical Evolution and the Origin of Life." *Scientific American* (Sep): 70 - 86

---. 1980. Dickerson, Richard E. "Cytochrome C and the Evolution of Energy Metabolism." *Scientific American* (Mar): 136 - 153.

Ditfurth, Hoimar v. 1982. The Origins of Life. Evolution as Creation. San Francisco: Harper and Row. Tr. (Peter Heinegg from *Wir sind nicht nur von dieser Welt: Naturwissenschaft, Religion und die Zukunft der Menschen.* Hamburg: Hoffmann und Campe, 1981)

Dobson, Eric J. 1955. "Early Modern Standard English." *Transactions of the Phonological Society* 1955. Reprinted in Lass 1969, 419 - 440.

Douglas, Matthew M. 1981. "Thermoregulatory Significance of Thoracic Lobes in the Evolution of Insect Wings." *Science* 211 (2 Jan): 84.

Downing, P. B. 1958-9. "Subjunctive Conditionals, Time Order and Causation." *Proceedings of the Aristotelian Society.*

Dummett, Michael and Antony Flew. 1954. "Can an Effect Precede Its Cause?" *Proceedings of the Aristotelian Society* (Supplementary Volume):

Durbin, John. 1979. *Modern Algebra.* New York: John Wiley.

Eakin, Richard and Jean L. Brandenberger. 1981. "Unique Eye of Probable Evolutionary Significance." *Science* 211: 1189 - 1190.

Ehrlich, Paul R. 1986. *The Machinery of Nature*. New York: Simon and Schuster.

Eigen, Manfred, William Gardiner, Peter Schuster, and Ruthild Winkler-Oswatitsch. 1981. "The Origin of Genetic Information." *Scientific American* (Apr): 88 - 119.

Ereshefsky, Marc. 1994. "Some Problems with the Linnaean Hierarchy." *Philosophy of Science* 61: 186-195.

Ettinger, Lia, Eva Jablonka, and Peter McLaughlin. 1990. "On the Adaptations of Organisms and the Fitness of Types." *Philosophy of Science,* 57: 499 - 513.

Euclid. 1956. *The Thirteen Books of the Elements* (2nd unabridged ed). Tr. Thomas L. Heath. New York: Dover.

Faber, Roger J. 1976. "Re-Encountering a Counter-Intuitive Probability." *Philosophy of Science*, 43: 283 - 285.

Fast, Julius 1970. *Body Language*. New York: Simon and Schuster.

Feigl, H., Sellars, W. and Lehrer, K., eds. 1972. *New Readings in Philosophical Analysis.* New York: Appleton-Century-Crofts.

Ferriere, Regis and Richard E. Michod. 1996. "The Evolution of Cooperation in Spatially Heterogeneous Populations." The American Naturalist 147: 692-717.

Feynman, Richard P. 1965. *The Character of Physical Law*. Cambridge, Mass.: The MIT Press.

---. 1985. *QED. The Strange Theory of Light and Matter*. Princeton, New Jersey: Princeton University Press.

"Is Darwin Obsolete." 1991. ("Firing Line" Television Series, 4 June 1991, No. 900). Columbia, SC: Southern Educational Communications Association.

Fitch, Walter M. 1982. "The Challenges to Darwinism Since the Last Centennial and the Impact of Molecular Studies." *Evolution* 36: 1133 - 1143.

Fox, L. R. and P. A. Morrow. 1981. "Specialization: Species Property or Local Phenomenon?" *Science* 211: 887 - 893.

Frege, Gottlob 1891. "Funktion und Begriff." In Frege 1994: 18 - 39.

Frege, Gottlob. ($^3$1993) *Logische Untersuchungen.* Göttingen: Vanderhoeck & Ruprecht.

---. $^7$1994 *Funktion, Begriff, Bedeutung. Fünf logische Studien.* Göttingen: Vanderhoeck & Ruprecht.

Frenzel, Louis E., Jr. 1987. *Crash Course in Artificial Intelligence and Expert Systems.* Indianapolis: Howard W. Sams.

Friedman, Michael. 1992. *Kant and the Exact Sciences.* Cambridge, Mass.: Harvard University Press.

Frisch, Karl von. 1965. *Tanzsprache und Orientierung der Bienen.* Berlin: Springer.

Frisch, Rose E. 1988. "Fatness and Fertility." *Scientific American* (Mar): 88 - 95.

Fromm, Erich. 1960. *Wege aus einer kranken Gesellschaft. Eine sozialpsychologische Untersuchung.* (Tr. Liselotte and Ernst Mickel). Frankfurt/M: Ullstein Materialien.

Futuyma, Douglas J. 1983. *Science on Trial. The Case for Evolution.* New York: Pantheon Books.

---. 1986. *Evolutionary Biology* (2nd Ed.). Sunderland, Mass.: Sinauer Associates.

Galanter, Marc. 1992. "Righting Old Wrongs." *Report from the Institute for Philosophy and Public Policy* 12 (Spring/Summer): 13 - 16.

Gamlin, Linda and Gail Vines. 1987. *The Evolution of Life.* New York: Oxford U. Press.

Gibbard, Allan 1990. *Wise Choices, Apt Feelings. A Theory of Normative Judgment.* Cambridge, Mass.: Harvard University Press.

Gilbert, W.S. 1878. "H.M.S. Pinafore, or the Lass that Loved a Sailor" (libretti). Reprinted in L.B. Lubin (ill.) *Gilbert without Sullivan* 1981. New York: Viking.

Gillies, Donald A. 1988. "Induction and Probability." In Parkinson, G.H.R., ed. *The Handbook of Western Philosophy*. New York: Macmillan.

Glance, Natalie S. and Bernardo A. Huberman. 1994. "The Dynamics of Social Dilemmas." *Scientific American* (Mar): 76 - 81.

Gleick, James. 1987. *Chaos. Making a New Science*. New York: Viking.

Gödel, Kurt. 1972. "Russell's Mathematical Logic." In Pears (ed.) 1972, 192 - 226.

Goethe, Johann Wolfgang von. 1790. *Die Metamorphose der Pflanzen*. In Bertha Müller (tr). *Goethe's Botanical Writings*. Woodbridge, Conn.: Ox Bow Press.

Goldberg, Samuel. 1976. "Copi's Conditional Probability Problem." *Philosophy of Science*, 43: 286 - 289.

Goldston, Linda 1996. "Mushrooms are deadly for family." *San Jose Mercury News*, 7 Feb. 1996, 4.

Goodman, Nelson. $^{2}$1976. *Languages of Art. An Approach to a Theory of Symbols* (rev. ed.) Indianapolis: Cambridge.

Gorczynski, R.M. and E.J. Steele. 1981. "Simultaneous yet Independent Inheritance of Somatically Acquired Inheritance of Two Distinct H-2 Antigenic Haplotype Determinants in Mice." *Nature* 289: 678-681.

Gore, Rick. 1996. "Neandertals." *National Geographic* (Jan): 2 - 35.

Gotthelf, Allan and James G. Lennox (eds). 1987. *Philosophical Issues in Aristotle's Biology*. Cambridge University Press.

Gould, Stephen Jay. 1977a. *Ever Since Darwin. Reflections in Natural History*. New York: Norton.

---. 1977b. *Ontogeny and Phylogeny*. Cambridge: Belknap.

---. 1980. *The Panda's Thumb. More Reflections in Natural History*. New York: Norton.

---. 1981. *The Mismeasure of Man*. New York: Norton.

---. 1983. *Hen's Teeth and Horse's Toes*. New York: Norton.

---. 1985. *The Flamingo's Smile*. New York: Norton.

---. 1989. *Wonderful Life*: The Burgess Shale and the Nature of History. New York: Norton.

---. 1991. *Bully for Brontosaurus*. New York: Norton.

---. 1993. *Eight Little Piggies*. New York: Norton.

---. 1994. “The Evolution of Life on the Earth.” *Scientific American* (Oct.): 85 - 91.

--- 1995. *Dinosaur in a Haystack. Reflections in Natural History*. New York: Harmony Books.

---. 1996. "Foreword" to Burkhardt 1996: ix - xx.

--- and R. Lewontin. 1979. “The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme.” *Proceedings of the Royal Society*, vol. B205, pp. 581-98. Reprinted in Sober 1984, pp. 252 - 270.

--- and Niles Eldredge. 1993. “Punctuated Equilibrium Comes of Age.” *Nature* 366: 223 - 27.

Greene, John C. 1974. “The History of Science and the History of Linguistics.” In Hymes 1974, 487 - 501.

Greenspan, Ralph J. 1995. “Understanding the Genetic Construction of Behavior.” *Scientific American* (Apr): 72 - 105.

Gribbin, John. 1984. *Schrödinger's Cat. Quantum Physics and Reality*. New York: Bantam.

Groves, David I., John S. R. Dunlop, and Roger Buick 1981. “Early Signs of Life.” *Scientific American* (Oct): 64 - 73.

Hacking, I. 1982. "Experimentation and Scientific Realism." *Philosophical Topics* 13: 71 - 87.

Hamilton, Edith and Huntington Cairns (eds) 1961. *The Collected Dialogues of Plato*. (Bollingen Series LXXI). Princeton: Princeton University Press.

Hamilton, W. D. 1964. “The Genetical Evolution of Social Behaviour.” *Journal of Theoretical Biology* 7: 1 - 32.

Hampe, Michael and S. R. Morgan. 1988. “Two consequences of Richard Dawkins' view of genes and organisms.” *Studies in the History and Philosophy of Science* 19: 119 - 138.

Haszprunar, Gerhard. 1994. “Ursprung und Stabilität tierischer Baupläne.” In Wieser (ed) 1994: 129 - 154.

Hawkins, Gordon and Franklin E. Zimring. 1988. *Pornography in a Free Society*. Cambridge: Cambridge University Press.

Heidegger, Martin. (1927). *Sein und Zeit.* Tübingen: Max Niemeyer Verlag, 1993.

---. (1927) *Being and Time*. Tr. John Macquarrie and Edward Robinson. New York: Harper and Row, 1962.

Hemleben, Johannes. 1968. *Darwin.* Hamburg: Rowohlt Taschenbuch Verlag.

Hempel, Carl G. 1960. "Inductive Inconsistencies." *Synthese*. Reprinted in Feigl *et al*. 1972.

Herstein, I. N. $^{2}$1975. *Topics in Algebra.* New York: John Wiley.

Hill, J. 1967. "The Environmental Induction of Heritable Changes in *Nicotiana rustica* Parental and Selection Lines." *Genetics* 55: 735-754.

Hobbes, Thomas. (1651). *Leviathan.* Harmonsdworth: Penguin, 1968.

Hofstadter, Douglas R. 1979. *Gödel, Escher, Bach: An Eternal Golden Braid.* New York: Vintage Books.

Holsinger, Kent E. 1984. "The Nature of Biological Species." *Philosophy of Science* 51: 293 - 307.

Horan, Barbara L. 1994. "The Statistical Character of Evolutionary Theory." *Philosophy of Science* 61: 76-95.

Horgan, John. 1995. "The Struggle Within." *Scientific American* (Jun): 16-18.

Hudlin, Charles W. 1993. *Moral Issues in Philosophy.* Dubuque, Iowa: Kendall/Hunt.

Hull, David. 1978. "A Matter of Individuality." *Philosophy of Science* 45: 335-360.

Hume, David. 1739. *A Treatise of Human Nature*. Oxford: Oxford University Press, 1975.

---. 1777. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press, 1975.

Hymes, Dell 1974. "Introduction: Traditions and Paradigms." In Hymes (ed) 1974: 1 - 38.

Hymes, Dell (ed) 1974. *Studies in the History of Linguistics. Traditions and Paradigms.* Bloomington and London: Indiana University Press.

Irving, David. 1963. *The Destruction of Dresden.* New York: MacMillan.

Jacobs, William P. 1994. "Caulerpa." *Scientific American* (Jun): 100-105.

Jakobson, Roman 1928. "The Concept of the Sound Law and the Teleological Criterion" (English translation of the Czech original). *Casopis pro moderni filologii*, XIV (March). Reprinted in Roman Jakobson 1961. *Selected Writings*, I, 1 - 2. The Hague: Mouton. Reprinted in Roger Lass (ed) 1969, 8 - 9.

--- 1932. "Phonetics and Phonology." First published in *Ottuv slovnik naucný* (Czech Encyclopædia). Prague. Reprinted in Roman Jakobson 1961. *Selected Writings*, I, 231-233. The Hague: Moulton. Reprinted in Roger Lass (ed) 1969, 6 - 7.

James, William. 1958. *The Varieties of Religious Experience. A Study in Human Nature (being the Gifford Lectures on Natural Religion delivered at Edinburgh in 1901 - 1902).* New York: New American Library.

Kac, Mark. 1964. "Probability." *Scientific American* (Sep): 92 - 111.

Kant, Immanuel. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels oder Versuch von der Verfassung und dem mechanischen Ursprunge des ganzen Weltgebäudes, nach Newtonschen Grundsätzen abgehandelt.* (Akademie-Textausgabe. Band I. Vorkritische Schriften I, 1747 - 1756). Berlin: Walter de Gruyter, 1968: 215 - 368.

Karthaus, Ulrich von (Hrg.) 1977. *Impressionismus, Symbolismus und Jugendstil* (Serie: Otto F. Best und Hans-Jürgen Schmitt (Hrg.) *Die deutsche Literatur. Ein Abriß in Text und Darstellung*, Band 13). Stuttgart: Phillipp Reclam.

Kastovsky, Dieter and Geri Bauer (eds) 1988. *Luick Revisited.* Tübingen: Narr.

Keegan, John. 1994. *A History of Warfare*. New York: Alfred Knopf.

Ker, R. F. and M. B. Bennett, S. R. Bibby, R. C. Kester, and R. McN. Alexander. 1987. "The spring in the arch of the human foot." *Nature* 325 (8 Jan): 147 - 149.

Kim, Jaegwon. 1978. "Supervenience and Nomological Incommensurables." *American Philosophical Quarterly* 15: 149 - 156.

---. 1984. "Concepts of Supervenience." *Philosophy and Phenomenological Research* 45: 153 - 176.

---. 1990. "Supervenience as a Philosophical Concept." *Metaphilosophy* 21: 1 - 27.

Kimura, Motoo. 1979. "The 'Neutral' Theory of Molecular Evolution." *Scientific American* (Nov): 98 - 129.

King, Stephen 1977. *The Shining*. New York: Signet.

---. 1985. "The Mist." In *Skeleton Crew*. New York: Signet.

Kipling, Rudyard 1890. "The Ballad of East and West." Reprinted in Rewey Belle Inglis and Josephine Spear 1958. *Adventures in English Literature*. New York: Harcourt, Brace & World, 547 - 550.

Kitchener, Andrew. 1987. "Scientific correspondence: Function of Claws' claws." *Nature* 325 (8 Jan): 114.

Kitcher, Philip. 1982. *Abusing Science. The Case Against Creationism.* Cambridge, Mass.: The MIT Press.

---. 1984. "Species." *Philosophy of Science,* 51: 308 - 333.

---. 1984b. "Against the Monism of the Moment: A Reply to Elliott Sober." *Philosophy of Science* 51: 616 - 630.

Kitts, David B. and David J. Kitts. 1979. "Biological Species as Natural Kinds." *Philosophy of Science*, 46: 613 - 622.

Klenk, Virginia. 1983. *Understanding Symbolic Logic.* New York: Prentice-Hall.

Knipe, David M. 1991. *Hinduism. Experiments in the Sacred.* New York: Harper Collins.

Koerner, E. F. K. 1978. "Four Types of History Writing in Linguistics." In E. F. K. Koerner 1978. *Towards a Historiography of Linguistics. Selected Essays.* Amsterdam: John Benjamins B.V., 55 - 62.

Koestler, Arthur. 1967. *The Ghost in the Machine*. New York: Macmillan.

Koyré, Alexandre. 1957. *From Closed World to Infinite Universe*. (Publications of the Institute of the History of Medicine, The Johns Hopkins University. Third Series: The Hideyo Noguchi Lectures, Vol. VII.) Baltimore and London: The Johns Hopkins University Press.

Kuhn, Thomas. $^{2}$1970. *Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lachterman, David R. 1989. *The Ethics of Geometry. A Genealogy of Modernity.* New York: Routledge.

Langer, Suzanne. 1957. *Philosophy in a New Key*. Cambridge: Harvard University Press.

---. $^{3}$1967. *An Introduction to Symbolic Logic*. New York: Dover.

Lass, Roger 1976. *English Phonology and Phonological Theory. Synchronic and Diachronic Studies*. Cambridge: Cambridge Univ. Pr.

--- 1986. "Conventionalism, invention, and 'historical reality.' Some reflections on method." *Diachronica* 3: 15 - 41.

--- 1987. *The Shape of English: Structure and History*. London: Dent.

--- 1988. "Vowel shifts, great and otherwise: remarks on Stockwell and Minkova." In Kastovsky and Bauer 1988: 395-410.

--- 1989. "How early does English get 'Modern'? Or, what happens if you listen to orthoepists and not to historians." *Diachronica* 6: 75 - 110.

--- 1992. "What, if anything, was the Great Vowel Shift?" In Risannen, Matti, Ossi Ihalainen, Terrtu Newalainen, Irma Taavitsainen (eds), *History of Englishes. New Methods and Interpretations in Historical Linguistics*, Berlin and New York: Mouton de Gruyter (Topics in English Linguistics, 10), 144-55.

Lass, Roger (ed) 1969. *Approaches to English Historical Linguistics. An Anthology*. New York: Holt, Rinehart and Winston.

Laudan, Larry. *Science and Relativism. Some Key Controversies in the Philosophy of Science*. Chicago: University of Chicago Press.

Leith, Dick 1983. *A Social History of English*. London and New York: Routledge.

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: John Wiley.

Lennox, James G. 1987. "Kinds, forms of kinds, and the more and the less in Aristotle's Biology." In Gotthelf and Lennox (1987): 339 - 359.

Lenzer, Gertrud (ed). 1975. *Auguste Comte and Positivism. The Essential Writings*. New York: Harper Torchbooks.

Levins, Richard. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54, no. 4: 421 - 431. Reprinted in Sober 1984c: 18 - 27.

Lewontin, Richard C. 1974. "The Structure of Evolutionary Genetics" in *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press: 3-18. Excerpted in Sober 1984c: 3 - 13.

--- 1978. "Adaptation." *Scientific American* (Sep): 212 - 230.

Lightner, Theodore. 1975. "The Role of Derivational Morphology in Generative Grammar." *Language* 51: 617 - 638.

Livingston, James C. $^2$1993. *Anatomy of the Sacred. An Introduction to Religion.* New York: Macmillan.

Lloyd, Elisabeth A. "A Semantic Approach to the Structure of Population Genetics." *Philosophy of Science*, 51: 242 - 264.

Lovejoy, C. Owen. 1981. "The Origin of Man." *Science* 211: 342 - 350.

Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge University Press.

---. 1977. *Semantics One*. Cambridge University Press.

---. 1977. *Semantics Two*. Cambridge University Press.

---. 1981. *Language and Linguistics*. Cambridge University Press.

Mackie, J. L. 1978. "The Law of the Jungle: Moral Alternatives and the Principles of Evolution." *Philosophy* 53: 455 - 64.

Maher, J. Peter, Allan R. Bomhard, and E. F. Konrad Koerner. 1982. *Papers from the 3rd International Conference on Historical Linguistics*. (Amsterdam Studies in the Theory and History of Linguistic Science. Series IV - Current Issues in Linguistic Theory, v. 13). Amsterdam: John Benjamins B. V.

Malthus, Thomas Robert. (1798). *An Essay on the Principle of Population.* Cambridge: Cambridge University Press, 1992.

---. (1820). *Principles of Political Economy.* Cambridge: Cambridge University Press, 1990.

Margulis, Lynn. 1971. "Symbiosis and Evolution." *Scientific American* (Aug): 48 - 57.

---. 1994. "Sex, Death and Kefir." *Scientific American* 271: 96.

Marshall, John C. 1987. "First squeaks of speech?" *Nature* 325 (15 Jan): 196.

Martin, Robert M. 1992. *There are Two Errors in the Title of the Book: A Sourcebook of Philosophical Puzzles, Paradoxes and Problems.* Peterborough, Ontario and Lewiston, NY: Broadview Press.

Martinet, André. (1981) *Sprachökonomie und Lautwandel. Eine Abhandlung über die diachronische Phonologie.* (tr. Claudia Fuchs). Stuttgart: Ernst Klett. (Originally published as *Économie des changements phonétiques*. Bern: A. Francke, 1955.)

Matthews, Warren. 1991. *World Religions.* St. Paul, Minn.: West Publishing.

May, Robert M. 1978. “The Evolution of Ecological Systems.” *Scientific American* (Sep): 160 - 175.

Maynard Smith, John 1970. "Time in the Evolutionary Process." *Studium Generale* 23: 266-272.

--- 1978. “The Evolution of Behavior.” *Scientific American* (Sep): 176 - 192.

--- and G. R. Price. 1973. “The Logic of Animal Conflicts.” *Animal Behaviour* 24: 159 - 175.

Mayr, Ernst. 1940. "Speciation phenomena in birds." *American Naturalist* 74: 249 - 278.

---. (1963). “Species Concepts and Their Applications.” *In Populations, Species, and Evolution* (Ch. 2). Cambridge, Mass.: Harvard University Press. Reprinted in Sober 1984c: 531 - 540.

---. 1964. “Introduction.” In Darwin 1964/(1859): vii - xxvii.

---. 1969. *Principles of Systematic Zoology*. New York: McGraw-Hill.

---. (1975). “Typological versus Population Thinking.” *Evolution and the Diversity of Life*. Cambridge, Mass.: Harvard University Press. Reprinted in Sober 1984c: 14 - 17.

---. 1978. “Evolution.” *Scientific American* (Sep): 46 - 55.

McKinney, Michael L. 1987. “Taxonomic selectivity and continous variation in mass and background extinctions of marine taxa.” *Nature* 325 (8 Jan): 143 - 145.

McLaughlin, William L. 1994. “Resolving Zeno’s Paradoxes.” *Scientific American* (Nov): 84 - 89.

McMahon, Thomas. 1987. “The spring in the human foot.” *Nature* 325 (8 Jan): 108 - 109.

McShea, Daniel W. 1991a. "Complexity and Evolution: What Everybody Knows." *Biology and Philosophy* 6: 303-324

---. 1994. "Mechanisms of Large-Scale Evolutionary Trends." *Evolution* 48(6): 1747- 1763.

Midgley, Mary. 1985. *Evolution as a Religion*. London: Methuen & Co.

Mills, Susan and John Beatty. 1979. "The Propensity Interpretation of Fitness," *Philosophy of Science* 46: 263 - 286. Reprinted in Sober 1984: 36 - 57.

Minkova, Donka 1991. "On leapfrogging in historical phonology." In Jaap van Marle (ed) *Historical Linguistics 1991: Papers from the 10th International Conference on Historical Linguistics*. (Amsterdam studies in the theory and history of linguistic science. Series IV, Current issues in linguistic theory, v. 107). Amsterdam and Philadelphia: John Benjamins Publishing Co., 211 - 228.

Minsky, Marvin. 1994. "Will Robots Inherit the Earth?" *Scientific American* (Apr): 108 - 113.

Moore, Edward F. 1964. "Mathematics in the Biological Sciences." *Scientific American* (Sep): 148 - 164.

Morgan, Ted 1993. *Wilderness at Dawn. The Settling of the North American Continent*. New York: Simon and Schuster.

Morris, Douglas W. 1986. "Proximate and Utimate Controls on Life-History Variation: The Evolution of Litter Size in White-Footed Mice (*Peromyscus leucopus*)." *Evolution* 40(1): 169 - 181.

Murray, James D. 1988. "How the Leopard gets its Spots." *Scientific American* (Mar): 80 - 87.

"New case parallels poisoning. Mushroom blamed in man's death." *Oakland Tribune*. 9 Feb. 1996, 2.

Newton, Isaac 1704. "On Fluxions." Tr. Evelyn Walker. In Smith 1959: 613 - 618.

Nowak, Martin A., Robert M. May, and Karl Sigmund. 1995. "The Arithmetic of Mutual Help." *Scientific American* (Jun): 76 - 83.

Orton, Harold and Martyn F. Wakelin (eds) 1967. *Survey of English dialects: The basic material, vol. 4. The Southern Counties. Parts 1 - 3*. Leeds: Arnold.

Otto, James and Towle, Albert. 1973. *Modern Biology*, p. 596. New York: Holt, Rinehart and Winston, 1973.

Owens, Joseph. 1981. *Aristotle. The Collected Papers of Joseph Owens*. John R. Catan, ed. Albany: State University of New York Press.

Pääbo, Svante. 1993. "Ancient DNA." *Scientific American* (Nov): 86 - 92.

Pears, D. F. (ed). 1972. *Bertrand Russell. A Collection of Critical Essays*. (Modern Studies in Philosophy Series) Garden City, NY: Anchor.

Peyser, Joan. 1987. *Leonard Bernstein. A Life*. New York: William Morrow.

Pfeiffer, John E. 1969. *The Emergence of Man*. New York: Harper and Row.

Pinker, Steven. 1994. *The Language Instinct.* New York: William Morrow.

Plato. *Meno.* Translated by W. K. C. Guthrie. In Hamilton and Cairns 1961: 353 - 384.

Platt, John, Heidi Weber, and Ho Mian Lian. 1984. *The New Englishes.* London: Routledge and Kegan Paul.

Popper, K.R. 1968. *Conjectures and Refutations: the Growth of Scientific Knowledge.* New York: Harper Torch Books.

Prior, John A. and Silberstein, Jack S. 1963. *Physical Diagnosis: The History and Examination of the Patient* (2nd ed.). Saint Louis, MO: C.V. Mosby Co.

Putnam, Hilary. 1990. *Realism with a Human Face.* Harvard University Press.

Quine, Willard van Orman. 1960. *Word and Object.* Cambridge, Mass.: MIT Press.

--- 1969. *Set Theory and its Logic* (rev. ed.). Cambridge, Mass.: Belknap.

Reichenbach, Hans. $^2$1949. *The Theory of Probability. An Inquiry into the Logical and Mathematical Foundations of the Calculus of Probability.* (Tr. Ernest H. Hutten and Maria Reichenbach) Berkeley and Los Angeles: University of California Press.

Restak, Richard M. *The Brain. The Last Frontier.* New York: Warner Books.

Ricardo, David (1817). *The Principles of Political Economy and Taxation.* London: Dent, 1973.

Ridley, Mark. 1985. *The Problems of Evolution.* Oxford: Oxford University Press.

Richards, Evelleen. 1994. "A Political Anatomy of Monsters, Hopeful and Otherwise." *Isis* 85: 377-411.

Rolnick, Philip A. 1993. *Analogical Possibilities. How Words Refer to God.* (American Academy of Religion Academy Series No. 81) Atlanta, GA: Scholars Press.

Rorty, Richard (ed) 1967. *The Linguistic Turn. Recent Essays in Philosophical Method.* Chicago and London: University of Chicago Press.

Rose, Lynn E. 1972. "Countering a Counter-Intuitive Probability." *Philosophy of Science*: 523 - 524.

Rosenberg, Alexander. 1985. *The Structure of Biological Science.* Cambridge University Press.

---. 1993. "Hume and the Philosophy of Science." In David Fate Norton (ed) 1993. *The Cambridge Companion to Hume*. Cambridge: Cambridge University Press: 64 - 89.

---. 1994. *Instrumental Biology or the Disunity of Science*. Chicago: University of Chicago Press.

Ross, Alan S. C. 1962. "U and Non-U: An Essay in Sociological Linguistics." In Max Black (ed) *The Importance of Language*. 1962. Ithaca, NY: Cornell University Press, 91 - 106. First published Nancy Mitford (ed) 1956. *Noblesse Oblige*. London: Hamish Hamilton, as an edited version of "Linguistic class-indicators in present-day English." *Neuphilologische Mitteilungen* 1954. Helsinki.

Ross, Sheldon. 1976. *A First Course in Probability*. New York and London: Macmillan and Collier Macmillan.

Ruse, Michael. 1993. *The Darwinian Paradigm: Essays on its History, Philosophy and Religious Implications*. London and New York: Routledge.

---. 1995. *Evolutionary Naturalism. Selected Essays*. London: Routledge.

Ruse, Michael (ed.). 1989. *What the Philosophy of Biology Is. Essays Dedicated to David Hull*. (Nijhoff International Philosophy Series, v. 32). Dordrecht: Kluwer Academic.

Russell, Bertrand. (1927). *An Outline of Philosophy*. New York: New American Library, 1960.

--- 1959. *Wisdom of the West*. London: Rathbone Books.

Ryle, Gilbert. 1931. "Systematically Misleading Expressions." *Proceedings of the Aristotelian Society*, XXXII (1931 - 32): 139 - 70. Reprinted in Rorty (ed) 1967, 85 - 100.

Sacks, Oliver. 1985. *The Man who Mistook his Wife for a Hat*. New York: Harper Collins.

Schneider, G. Michael, Steven W. Weingart, and David M. Perlman. 1978. *An Introduction to Programming and Problem Solving with Pascal*. New York: John Wiley.

Schopf, J. William 1978. "Evolution of the First Cells." *Scientific American* (Sep): 110 - 138.

Searle, John R. 1992. *The Rediscovery of the Mind*. Cambridge, Mass.: The MIT Press.

Seppy, Tom 1994. "Hail Columbus." *Reader's Digest* (10): 88.

Sigurbjörnsson, Björn. 1971. "Induced Mutations in Plants." *Scientific American* (Jan): 86 - 95.

Simons, Elwyn L. and Peter C. Ettel. 1970. "Gigantopithecus." *Scientific American* (Jan): 76 - 85.

Simpson, George Gaylord. (1949.) *The Meaning of Evolution: A Study of the History of Life and Its Significance for Man* (revised ed.). New Haven, Conn.: Yale University Press, 1967.

---. 1961. *Principles of Animal Taxonomy.* New York: Columbia University Press.

Slurink, Pouwel 1996. "Back to Roy Wood Sellars: Why His Evolutionary Naturalism Is Still Worthwhile." *Journal of the History of Philosophy* 34: 425 - 449.

Smart, Ninian and Richard D. Hecht (eds). 1982. *Sacred Texts of the World. A Universal Anthology.* New York: Crossroad.

Smith, Adam. 1759. *The Theory of Moral Sentiments.* Indianapolis, Ind.: Liberty Press, 1982.

---. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations.* Oxford: Oxford University Press, 1976.

Smith, David E. 1959. *A Source Book in Mathematics.* Mineola, NY: Dover.

Sober, Elliott. 1981. "Evolution, Population Thinking and Essentialism." *Philosophy of Science* 47: 350 - 383.

---. 1984a. "Common Cause Explanation." *Philosophy of Science* 51: 212 - 241.

---. 1984b. *The Nature of Selection. Evolutionary Theory in Philosophical Focus*, Cambridge, Mass.: The MIT Press.

---. (ed). 1984c. *Conceptual Issues in Evolutionary Biology.* Cambridge, Mass.: The MIT Press.

---. 1988. *Reconstructing the Past. Parsimony, Evolution, and Inference.* Cambridge, Mass.: The MIT Press.

---. 1993. *Philosophy of Biology.* Oxford: Oxford University Press.

Sorrel, Tom. 1986. *Hobbes.* London and New York: Routledge.

Stanford, P. Kyle. 1995. "For Pluralism and Against Realism about Species." *Philosophy of Science* 62: 70 - 91.

Stevin, Simon. (1634). On Decimal Fractions. Tr. Vera Sanford. In Smith 1959: 20 - 34.

Stockwell, Robert P. 1969. "Mirrors in the History of English Pronunciation." In Lass (ed.) 1969: 228 - 245.

Stockwell, Robert P. 1969. "On the Utility of an Overall Pattern in Historical English Phonology." *Proceedings of the Ninth International Congress of Linguistics* (1962). The Hague, 663 - 671. Reprinted in Lass 1969, 88 - 96.

--- and Donka Minkova 1988a. "The English Vowel Shift: problems of coherence and explanation." In Kastovsky and Bauer 1988: 355-94.

--- and --- 1988b. "A rejoinder to Lass." In Kastovsky and Bauer 1988: 411-17.

Strawson, P. F. 1966. *The Bounds of Sense. An Essay on Kant's Critique of Pure Reason*. London: Methuen.

Strossen, Nadine. 1987. *Defending Pornography. Free Speech, Sex, and the Fight for Women's Rights*. New York: Scribners.

Suppes, Patrick. 1984. *Probabilistic Metaphysics*. Oxford: Basil Blackwell.

Taylor, T. G. 1970. "How an Eggshell is Made." *Scientific American* (Mar): 88 - 97.

Taylor, William A. 1988. *What Every Engineer Should Know about Artificial Intelligence*. Cambridge, Mass.: The MIT Press.

Teilhard de Chardin, Pierre. 1956. *The Appearance of Man*. New York: Harper and Row. (Tr. J. M. Cohen from *L'apparition de l'homme*. Paris: Edition du Seuil.)

---. (1957). *The Vision of the Past*. New York: Harper and Row, 1966. (Tr. J. M. Cohen from *La vision du passé*. Paris: Edition du Seuil.)

Thoreau, Henry D. (1854). *Walden or, Life in the Woods*. New York: New American Library, 1960.

Tijan, Robert. 1995. "Molecular Machines that Control Genes." *Scientific American* (Feb): 54 - 61.

Todd, James T., Leonard S. Mark, Robert E. Shaw, and John B. Pittenger. 1980. "The Perception of Human Growth." *Scientific American* (Feb): 132 - 144.

Toulmin, Stephen. 1972. *Human Understanding*. Vol. 1. Princeton, N.J.: Princeton University Press.

Trinkaus, Eric and William W. Howells. 1979. "The Neanderthals." *Scientific American* (Dec): 118 - 133.

Valentine, James W. 1978. "Evolution of Multicelled Plants and Animals." *Scientific American* (Sep): 140 - 158.

van Lawick-Goodall, Jane. 1971. *In the Shadow of Man*. New York: Dell.

Wachbroit, Robert. 1994. "Normality as a Biological Concept." *Philosophy of Science* 61: 579 - 591.

Wagner, Günter P. 1994. "Der Dialog zwischen Evolutionsforschung und Computerwissenschaft." In Wieser (ed) 1994: 221 - 233.

Walker, Alan and Richard E. F. Leakey 1978. "The Hominids of East Turkana." *Scientific American* (Aug): 54 - 75.

Walzer, Michael. $^2$1992. *Just and Unjust Wars. A Moral Argument with Historical Illustrations*. New York: HarperCollins.

Washburn, Sherwood L. 1978. "The Evolution of Man." *Scientific American* (Sep): 194 - 208.

Wasserstrom, Richard A. 1977. "Racism and Sexism." 24 *UCLA Law Review* 581. Reprinted in Hudlin 1993: 21 - 42.

Waterlow, Sarah. 1974. "Backwards Causation and Continuing." *Mind*:

Weber, Marcel. 1996. "Fitness Made Physical: The Supervenience of Biological Concepts Revisited." *Philosophy of Science* 63: 411 - 431.

Weeks, David and Jamie James. 1995. *Eccentrics. A Study of Sanity and Strangeness*. New York: Villard Books.

Wells, Rulon 1974. "Phonemics in the Nineteenth Century, 1876 - 1900." In Hymes 1974: 434 - 453.

Whitehead, Alfred North 1964. *The Concept of Nature*. Cambridge: Cambridge University Press.

Wicksten, Mary K. 1980. "Decorator Crabs." *Scientific American* (Feb): 146 - 154.

Wieser, Wolfgang (ed). 1994. *Die Evolution der Evolutionstheorie. Von Darwin zur DNA*. Heidelberg: Spektrum Akademischer Verlag.

Williams, G. C. 1966. *Adaptation and Natural Selection.* Princeton: Princeton University Press.

Wilmot, William W. $^{3}$1987. *Dyadic Communication.* New York: Random House.

Wilson, Edward O. 1975. *Sociobiology* (abridged ed.). Cambridge, Mass.: Harvard University Press.

Winfrey, Arthur T. 1987. *The Timing of Biological Clocks*. (Scientific American Library 19) New York: Scientific American Books.

Wirth, Niklaus. 1986. *Algorithms and Data Structures*. Englewood Cliffs, N.J.: Prentice-Hall.

Woese, Carl R. 1981. "Archaebacteria." *Scientific American* (Jun): 98 - 125.

# Autobiographical Notes

From September 1994 through February 1995 I studied German at the *Internazionales Studienzentrum der Universität-Heidelberg*. Since October 1994 I have been enrolled at the university itself, with *Philosophie* and *Anglistik* as major subjects. Previous degrees include: B.A. in philosophy from Brandeis University (Waltham, Massachusetts), B.A. in mathematics (University of Colorado, Boulder), M.A. in philosophy (The Catholic University of America, Washington, D.C.).